

RESULT	1
LOCUS	BG551127
DEFINITION	BG551127 491 bp mRNA EST 09-APR-2001 S8d33f01.y1 Gm-cl074 Glycine max CDNA clone GENOME SYSTEMS CLONE ID: Gm-cl074-2162 5' smallar to TR:Q9SVI4 Q9SVI4 ES43 LIKE PROTEIN // mRNA sequence.
ACCESSION	BG551127
VERSION	BG551127.1 GI:13562907
KEYWORDS	EST.
SOURCE	soybean.
ORGANISM	Glycine max Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae Glycine.
REFERENCE	1 (bases 1 to 491)
AUTHORS	Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Treising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shiu, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R. Public Soybean EST project Unpublished (1999) Contact: Shoemaker R/Public Soybean EST project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or
TITLE	
JOURNAL	
COMMENT	



info@genomesystems.com web site: www.genomesystems.com  
High quality sequence stop: 421.

## FEATURES

Location/Qualifiers

1..491  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1074-2162"  
/clone\_1lb="Gm-c1074"  
/clone\_type="seedlings induced for HR (hypersensitive response)"  
/dev\_stage="9-11 day old"  
/lab\_host="DH10B"  
/note="Vector: Bluescript II SK(+); Site 1: EcoRI; Site 2: XhoI. The cDNA library was constructed from mRNA isolated from 9-11 day old seedlings that were induced for HR (hypersensitive response) by vacuum infiltrating plant tissue with *Pseudomonas syringae* pv. *glycinea* carrying the *avrB* gene (Genetics 141:1597-1604). Plant tissue (expanded unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36, and 53 hrs after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene Bluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into E. coli Electromax DH10B host cells. Plant care, inoculations, and library construction were performed by Steve Clough (Lila Vodkin lab, University of Illinois)."

## BASE COUNT

117 a 138 c 129 g 107 t

## ORIGIN

## Query Match

Best local Similarity 36.6%; Score 45; DB 11; Length 491;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaaatgtgagatgccttaaacacctgagcctaattg 47  
|||||  
Db 430 ggtattgtaaatgtgagatgccttaaacacctgagcctaattg 474

## RESULT 2

## LOCUS

## DEFINITION

BG551208 497 bp mRNA EST 09-APR-2001  
sad34e05.y1 Gm-c1074 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1074-2145 5' similar to TR:O5SV14 O9SV14 ES43 LIKE PROTEIN ; mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

1 (bases 1 to 497)  
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rutter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., Mccann, R., Waterston, R. and Wilson, R.

TITLE  
JOURNAL  
COMMENT  
Public Soybean EST Project  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project

## FEATURES

source

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estevatson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63114 For further information call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
High quality sequence stop: 380.

Location/Qualifiers

1..497  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1074-2145"  
/clone\_1lb="Gm-c1074"  
/clone\_type="seedlings induced for HR (hypersensitive response)"  
/dev\_stage="9-11 day old"  
/lab\_host="DH10B"  
/note="Vector: Bluescript II SK(+); Site 1: EcoRI; Site 2: XhoI. The cDNA library was constructed from mRNA isolated from 9-11 day old seedlings that were induced for HR (hypersensitive response) by vacuum infiltrating plant tissue with *Pseudomonas syringae* pv. *glycinea* carrying the *avrB* gene (Genetics 141:1597-1604). Plant tissue (expanded unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36, and 53 hrs after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene Bluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into E. coli Electromax DH10B host cells. Plant care, inoculations, and library construction were performed by Steve Clough (Lila Vodkin lab, University of Illinois)."

## BASE COUNT

119 a 140 c 129 g 109 t

## ORIGIN

## Query Match

Best local Similarity 35.8%; Score 44; DB 11; Length 497;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttattgtaaatgtgagatgccttaaacacctgagcctaattg 47  
|||||  
Db 431 gttattgtaaatgtgagatgccttaaacacctgagcctaattg 474

## RESULT 3

## LOCUS

## DEFINITION

AW761136 530 bp mRNA EST 21-NOV-2000  
sl63f12.y1 Gm-c1077 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1027-6024 5' similar to TR:O65462 O65462 RECEPTOR LIKE PROTEIN ; mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

REFERENCE  
1 (bases 1 to 530)  
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rutter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., Mccann, R., Waterston, R. and Wilson, R.







Oy 4 gtaattgtaaatgtgagatgccttacacccctgatacctaagg 47  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 566 GTATTGTAATGTGAGATGCTTACCAACCTGACTGACTAATG 523

RESULT 5  
 AM432561 540 bp mRNA EST 18-JUL-2000  
 LOCUS AM432561  
 DEFINITION sh75h06.y1 Gm-cl015 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl015-5652 5' similar to TR:065462 065462 RECEPTOR LIKE PROTEIN ; , mRNA sequence.  
 AM432561  
 ACCESSION AM432561.1 GI:6963868  
 VERSION EST.  
 KEYWORDS soybean.  
 SOURCE Glycine max  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 540)  
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Merritt,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Thelming,B., Allen,M., Bowers,I., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rletter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
 Insert length: 1008 Std Error: 0.00  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 380.  
 Location/Qualifiers  
 1..540  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl015-5652"  
 /clone\_lib="Gm-cl015"  
 /tissue\_type="Mature flowers, field grown plants"  
 /lab\_host="XL10-Gold"  
 /note="Vector: pBluescript II XR; Site1: EcoRI; Site2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."

BASE COUNT  
 ORIGIN

123 a 142 c 138 g 137 t

Query Match 23.6%; Score 29; DB 10; Length 540;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 gtaattgtaaatgtgagatgccttacaccc 32  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 468 GTATTGTAATGTGAGATGCTTACCAACC 496

RESULT 6  
 BE659470 765 bp mRNA EST 24-MAY-2001  
 LOCUS BE659470/c  
 DEFINITION GM700009B10H8 Gm-r1070 glycine max cDNA clone Gm-r1070-3447 3', mRNA sequence.  
 BE659470  
 ACCESSION BE659470.1 GI:9985362  
 VERSION EST.  
 KEYWORDS soybean.  
 SOURCE Glycine max  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 765)  
 AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Erpelding,J., Rapp,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.  
 TITLE A Functional Genomics Program for Soybean (NSF 9872565)  
 JOURNAL Unpublished (1999)  
 COMMENT Other ESTs: AM432561 corresponding to Gm-cl015-5652 (5')  
 Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)  
 Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics  
 University of Illinois  
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
 Tel: (217) 244-6147  
 Fax: (217) 333-4582  
 Email: l-vodkin@uiuc.edu  
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.  
 Location/Qualifiers  
 1..765  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="Gm-r1070-3447"  
 /clone\_lib="Gm-r1070"  
 /note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/research/Projects/Soybean/Index.html. Rerecking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT

227 a 167 c 150 g 197 t 24 others



## ORIGIN

Query Match 21.1%: Score 26; DB 10; Length 765;  
 Best Local Similarity 100.0%: Pred. No. 0.0011;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 gcaatgcagagcgcactgactgt 74  
 ||||||||||||||||||  
 Db 497 GCATGCGAGGCGCTGACTGCTGT 472

RESULT 7  
 A0451805/c 349 bp DNA GSS 21-APR-1999  
 LOCUS HS.5183.A1.B08.T7A.RPCR-11 Human Male BAC Library Homo sapiens  
 DEFINITION genomic clone Plate=759 Col-15 Row-C, DNA sequence.  
 ACCESSION A0451805  
 VERSION A0451805.1 GI:4592975  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 349)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

JOURNAL MEDLINE  
 COMMENT 99380589  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Piere de Jong  
 (piere@dejong.med.bufileo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.bufileo.edu/ordering\_bac.htm)  
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
 http://www.htsc.washington.edu  
 Plate: 759 row: C column: 15  
 Seq primer: 17  
 Class: BAC ends  
 High quality sequence stop: 349.

## FEATURES

source  
 1. 349  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=759 Col=15 Row=C"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACe3.6 vector at EcoRI sites"  
 BASE COUNT 83 a 79 c 86 g 101 t  
 ORIGIN

Query Match 17.1%: Score 21; DB 13; Length 349;  
 Best Local Similarity 100.0%: Pred. No. 0.53; Length 349;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ttatctcaattatccaca 118  
 ||||||||||||||||||  
 Db 69 TTATCTCAATTATTCACA 49

## RESULT 8

BE999004 457 bp mRNA EST 06-OCT-2000  
 DEFINITION EST430727 GVSN Medicago truncatula cDNA clone pgvsn-13016, mRNA  
 sequence.  
 ACCESSION BE999004  
 VERSION BE999004.1 GI:10699280  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 457)  
 Fedorova,M., Pierson,B.L., Samac,D.A., Gantl,J.S., Vance,C.P., Town  
 C.D., Bowman,C.L., Craven,M.B., Cho,J., and Fraser,C.M.  
 ESTs from senescent nodules of Medicago truncatula  
 Unpublished (2000)  
 CONTACT: Caroll P. Vance  
 Department of Agronomy and Plant Genetics  
 University of Minnesota  
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
 Tel: 612 625 5715  
 Fax: 651-649-5058  
 Email: Vance004@maroon.tc.umn.edu  
 University of Minnesota name: M272982e TIGR sequence name:  
 MTKAX20TK More information is available at:  
 http://chryslie.tamu.edu/medicago  
 Seq primer: SKmod (CTA GAA CTA gtc gat CC).

## FEATURES

source  
 1. 457  
 Location/Qualifiers  
 /organism="Medicago truncatula"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="pgvsn-13016"  
 /clone\_lib="GVSN"  
 /rissue\_type="senescent root nodules"  
 /dev\_stage="mixture of effective nodules from 40 day old  
 plants harvested 36 hours post shoot removal and nodules  
 collected from 2 month old plants at mid-pod stage"  
 /lab\_host="E. coli strain SOLR"  
 /note="Vector: pBluescript SK +/- Site.1: EcoRI; Site.2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from the  
 mixture of effective nodules of 40 day old plants  
 harvested 36 hours post shoot removal and nodules  
 collected from 2 month old plants at mid-pod stage. The  
 cDNA was directionally ligated into the Uni-ZAP XR vector  
 from stratagene and packaged using Gigapack III GOLD  
 packaging extracts. Plasmids containing cDNA inserts were  
 excised from the recombinant lambda-ZAP phage using  
 Ex-assist helper phage and propagated in SOLR cells."  
 BASE COUNT 126 a 90 c 109 g 132 t  
 ORIGIN

Query Match 17.1%: Score 21; DB 11; Length 457;  
 Best Local Similarity 100.0%: Pred. No. 0.54;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gtaattgaatgtgagatgcc 24  
 ||||||||||||||||||  
 Db 416 GTATTGTAAATGTGAGATGCC 436

RESULT 9  
 AM586761 514 bp mRNA EST 07-SEP-2000  
 LOCUS EST318384 MRAM Medicago truncatula/61omus versiforme mixed EST  
 DEFINITION library cDNA clone pMHAM-55612, mRNA sequence.  
 ACCESSION AM586761  
 VERSION AM586761.1 GI:7266275



KEYWORDS EST.  
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.  
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library.  
REFERENCE 1 (bases 1 to 514)  
AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
TITLE ESTs from roots of Medicago truncatula after colonization with Glomus versiforme  
JOURNAL Unpublished (2000)  
COMMENT Contact: Maria J. Harrison  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73401, USA  
Tel: 580-223-5810  
Fax: 580-221-7380  
Email: mjharrison@noble.org  
Noble EST name: N25126  
TIGR sequence name: WTD4V42TK  
More information is available at: 'http://chryslie.tamu.edu/medicago/'  
Seq primer: SKmod (CTA gAA CTA gAT CC).  
Location/Qualifiers  
1. .514  
/organism="Medicago truncatula/Glomus versiforme mixed EST library"  
/cultivar="Medicago truncatula genotype A17"  
/db\_xref="taxon:119092"  
/clone\_1db="MHAM-55G12"  
/clone\_1lb="MHAM"  
/tissue\_type="roots colonized with Glomus versiforme"  
/dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."  
/lab\_host="E. coli strain XLOLR"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the Uniap XR vector from stragene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."  
BASE COUNT 135 a 101 c 127 g 151 t  
ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 514;  
Best Local Similarity 100.0%; Pred. No. 0.54;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttatgtaaatgtgagatgcc 24  
|||||  
Db 415 GATTGTAATGTGAGATGCC 435

RESULT 10  
BF650924 539 bp mRNA EST 20-DEC-2000  
LOCUS NF098E01EC1F1005 Elicited cell culture Medicago truncatula cDNA  
DEFINITION clone NF098E01EC 5', mRNA sequence.  
ACCESSION BF650924  
VERSION BF650924.1 GI:11916054  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
REFERENCE 1 (bases 1 to 539)  
AUTHORS Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,

TITLE Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research  
JOURNAL Unpublished (2000)  
COMMENT Contact: Dixon RA  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7302  
Fax: 580 221 7380  
Email: radixon@noble.org  
Insert Length: 539 Std Error: 0.00  
Insert Length: 539 row: E column: 01  
Plate: 098  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1. .539  
/organism="Medicago truncatula"  
/db\_xref="taxon:13880"  
/clone\_1db="NF098E01EC"  
/clone\_1lb="Elicited cell culture"  
/tissue\_type="Cell cultures derived from root tissues"  
/dev\_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"  
/note="Vector: Lambda Zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5, 1, 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."  
BASE COUNT 146 a 108 c 125 g 157 t 3 others  
ORIGIN

Query Match 17.1%; Score 21; DB 11; Length 539;  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttatgtaaatgtgagatgcc 24  
|||||  
Db 443 GATTGTAATGTGAGATGCC 463

RESULT 11  
BE124753 574 bp mRNA EST 07-SEP-2000  
LOCUS EST393788 GVN Medicago truncatula cDNA clone gVN-67F7, mRNA  
DEFINITION sequence.  
ACCESSION BE124753  
VERSION BE124753.1 GI:8529310  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
REFERENCE 1 (bases 1 to 574)  
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula  
JOURNAL Unpublished (2000)  
COMMENT Contact: Carroll P. Vance  
Department of Agronomy and Plant Genetics  
University of Minnesota  
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
Tel: 612 625 5715  
Fax: 651-649-5058  
Email: vance004@maroon.tc.umn.edu  
University of Minnesota name: N261622e  
TIGR sequence name: MTCC28TK  
More information is available at: 'http://chryslie.tamu.edu/medicago'



FEATURES  
source  
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).  
Location/Qualifiers  
1..574

/organism="Medicago truncatula"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="pGVN-67F7"  
/clone.lib="GVN"  
/tissue\_type="N2-fixing root nodules"  
/dev\_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"  
/lab\_host="E. coli strain XLOLR"  
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from Strategene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells."

BASE COUNT  
ORIGIN  
155 a 114 c 137 g 167 t 1 others

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gttatgtaaatgtgagatgcc 24  
|||||

Db 439 GTATTGTAATGTGAGATGCC 459

RESULT 12  
BE997991 586 bp mRNA EST 06-OCT-2000  
LOCUS EST429714 GVSN Medicago truncatula cDNA clone pGVSN-8F9, mRNA  
DEFINITION  
sequence.  
ACCESSION BE997991  
VERSION BE997991.1 GI:10698267  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
1 (bases 1 to 586)  
Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.  
ESTs from senescent nodules of Medicago truncatula  
Unpublished (2000)  
Contact: Carroll P. Vance  
Department of Agronomy and Plant Genetics  
University of Minnesota  
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
Tel: 612 625 5715  
Fax: 651-649-5058  
Email: vance004@maroon.tc.umn.edu  
University of Minnesota name: M271969e TIGR sequence name: MTKRKK297K More information is available at: http://chrysis.tamu.edu/medicago  
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).  
Location/Qualifiers  
1..586

FEATURES  
source  
/organism="Medicago truncatula"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="pGVSN-8F9"  
/clone.lib="GVSN"  
/tissue\_type="senescent root nodules"  
/dev\_stage="mixture of effective nodules from 40 day old

plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage"  
/lab\_host="E. coli strain SOLR"  
/note="Vector: pBluescript SK +/-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of effective nodules of 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage. The cDNA was directionally ligated into the Uni-ZAP XR vector from Strategene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT  
ORIGIN  
158 a 114 c 143 g 171 t

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gttatgtaaatgtgagatgcc 24  
|||||

Db 432 GTATTGTAATGTGAGATGCC 452

RESULT 13  
AM584723 635 bp mRNA EST 07-SEP-2000  
LOCUS N210838e MHAM Medicago truncatula/Glomus versiforme mixed EST  
DEFINITION  
library cDNA clone MHAM-7K19, mRNA sequence.  
ACCESSION AM584723  
VERSION AM584723.1 GI:7261777  
KEYWORDS EST.  
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.  
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library  
Eukaryota; mixed EST libraries.  
1 (bases 1 to 635)  
Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
ESTs from roots of Medicago truncatula after colonization with Glomus versiforme  
Unpublished (2000)  
Contact: Harrison M.J.  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73401  
Tel: 580-223-5810  
Fax: 580-223-7380  
Email: mjharrison@noble.org  
Other name: MHAM-7a-P10; Date: 3/14/00; Updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at: http://chrysis.tamu.edu/medicago".  
Seq primer: T3.  
Location/Qualifiers  
1..635

FEATURES  
source  
/organism="Medicago truncatula/Glomus versiforme mixed EST library"  
/cultivar="Medicago truncatula genotype A17"  
/db\_xref="taxon:119092"  
/clone="MHAM-7K19"  
/clone.lib="MHAM"  
/tissue\_type="roots colonized with Glomus versiforme"  
/dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."  
/lab\_host="E. coli strain XLOLR"  
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the UniZAP XR vector from Strategene and packaged using Gigapack III Gold packaging



extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 202 a 159 c 103 g 171 t

ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 635;  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttatgtaaatgtgagatgcc 24  
|||||

Db 468 GTATTGTAATGTGAGATGCC 448

RESULT 14  
AM584724/c 655 bp mRNA EST 07-SEP-2000  
LOCUS N210839e MHAM Medicago truncatula/Glomus versiforme mixed EST  
DEFINITION library cDNA clone MHAM-7K21, mRNA sequence.  
ACCESSION AM584724  
VERSION AM584724.1 GI:7261778  
KEYWORDS EST.  
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.  
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library.  
REFERENCE 1 (bases 1 to 655)  
AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
TITLE ESTs from roots of Medicago truncatula after colonization with Glomus versiforme  
JOURNAL Unpublished (2000)  
COMMENT Contact: Harrison M.J.  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73401  
Tel: 580-223-5810  
Fax: 580-223-7380  
Email: mharrison@noble.org  
Other name: MHAM-7a-F11; Date: 3/14/00; updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at 'http://chrystle.tamu.edu/medicago'.  
Seq primer: T3.

FEATURES

source

1..655  
Location/Qualifiers  
/organism="Medicago truncatula/Glomus versiforme mixed EST library"  
/cultivar="Medicago truncatula genotype A17"  
/db\_xref="taxon:119092"  
/clone="MHAM-7K21"  
/clone\_1id="MHAM"  
/tissue\_type="roots colonized with Glomus versiforme"  
/dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."  
/lab\_host="E. coli strain XL0LR"  
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the UniZap XR vector from Stragene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 207 a 163 c 108 g 177 t

ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 655;  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttatgtaaatgtgagatgcc 24  
|||||

Db 468 GTATTGTAATGTGAGATGCC 448

RESULT 15  
AM584948 682 bp mRNA EST 07-SEP-2000  
LOCUS N211110e MHAM Medicago truncatula/Glomus versiforme mixed EST  
DEFINITION library cDNA clone MHAM-7J12, mRNA sequence.  
ACCESSION AM584948  
VERSION AM584948.1 GI:7262002  
KEYWORDS EST.  
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.  
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library.  
REFERENCE 1 (bases 1 to 682)  
AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
TITLE ESTs from roots of Medicago truncatula after colonization with Glomus versiforme  
JOURNAL Unpublished (2000)  
COMMENT Contact: Harrison M.J.  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73401  
Tel: 580-223-5810  
Fax: 580-223-7380  
Email: mharrison@noble.org  
Other name: MHAM-7d-E06; Date: 3/14/00; updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at 'http://chrystle.tamu.edu/medicago'.  
Seq primer: T3.

FEATURES

source

1..682  
Location/Qualifiers  
/organism="Medicago truncatula/Glomus versiforme mixed EST library"  
/cultivar="Medicago truncatula genotype A17"  
/db\_xref="taxon:119092"  
/clone="MHAM-7J12"  
/clone\_1id="MHAM"  
/tissue\_type="roots colonized with Glomus versiforme"  
/dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."  
/lab\_host="E. coli strain XL0LR"  
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the UniZap XR vector from Stragene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 212 a 172 c 115 g 183 t

ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 682;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttatgtaaatgtgagatgcc 24  
|||||

Db 468 GTATTGTAATGTGAGATGCC 448

RESULT 16  
AO597145/c 444 bp DNA GSS 08-JUN-1999  
LOCUS HS-5240\_A1\_B11-SP6E RPCT-11 Human Male BAC Library Homo sapiens  
DEFINITION



genomic clone Plate-816 Col-21 Row-C, DNA sequence.

ACCESSION A0597145  
 VERSION A0597145.1 GI:5028357  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 444)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,U., Young,U., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 816 row: C column: 21  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 444.

FEATURES  
 source  
 Location/Qualifiers  
 1..444  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-816 Col-21 Row-C"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT 111 a 96 c 103 g 131 t 3 others

ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 444;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 titatctcaattatctc 115  
 ||||||||||||||||  
 Db 47 TTAATCTCAATTAATCTC 30

RESULT 17  
 BE653345 468 bp mRNA EST 06-SEP-2000  
 LOCUS UI-M-ALL-ah1-c-07-0-UI r1 NIH BMAP\_MCO\_N Mus musculus cDNA clone  
 DEFINITION UI-M-ALL-ah1-c-07-0-UI 5', mRNA sequence.  
 ACCESSION BE653345  
 VERSION BE653345.1 GI:9979245  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus. 1 (bases 1 to 468)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mestr@nmi.nih.gov  
 CDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 Location/Qualifiers  
 1..468  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-ALL-ah1-c-07-0-UI"  
 /clone\_lib="NIH\_BMAP\_MCO\_N"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The NIH\_BMAP\_MCO\_N library is a normalized library constructed from mouse cortex. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."

BASE COUNT 102 a 135 c 136 g 95 t

ORIGIN

Query Match 14.6%; Score 18; DB 10; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 gagggctgactgactg 73  
 ||||||||||||||||  
 Db 185 GAGGCTGCACTGACTG 202

RESULT 18  
 A0597171 487 bp DNA GSS 08-JUN-1999  
 LOCUS HS\_5240\_A1\_E12\_SPE RPCI-11 Human Male BAC Library Homo sapiens  
 DEFINITION genomic clone Plate-816 Col-23 Row-I, DNA sequence.  
 ACCESSION A0597171  
 VERSION A0597171.1 GI:5028383  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 487)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,U., Young,U., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618



Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Research Genetics (<http://www.htsc.washington.edu>). BAC end Web Server: <http://www.htsc.washington.edu>  
 Plate: 816 row: 1 column: 23  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 487.

## FEATURES

source

Location/Qualifiers

1..487

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=816 Col=23 Row=1"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

## BASE COUNT

114 a 111 c 114 g 140 t 8 others

## ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 487;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 tcttaattattccaca 118  
 ||||||||||||||||  
 Db 44 TCTTCATTATTCACA 27

RESULT 19  
 BH054207/c 533 bp DNA GSS 17-JUL-2001  
 LOCUS  
 DEFINITION  
 , DNA sequence.  
 BH054207  
 BH054207.1 GI:14849018

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 533)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Other\_GSSs: RPCI-24-337C12.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208

Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://www.tigr.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 337 row: C column: 12  
 Seq primer: SP6  
 Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..533

## BASE COUNT

154 a 100 c 81 g 198 t

## ORIGIN

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-337C12"  
 /clone\_lib="RPCI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

## BASE COUNT

154 a 100 c 81 g 198 t

## ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 533;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 taagttcacatgtaag 91  
 ||||||||||||||||  
 Db 266 TAGTTTCACATGTAGG 249

RESULT 20  
 A2290583/c 574 bp DNA GSS 27-JUL-2000  
 LOCUS  
 DEFINITION  
 , DNA sequence.  
 A2290583  
 A2290583.1 GI:9532369

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 574)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Other\_GSSs: RPCI-23-129H21.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208

Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Research Genetics (<http://www.htsc.washington.edu>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 129 row: H column: 21  
 Seq primer: T7  
 Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..574

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-129H21"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size



selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). \*

Query Match 14.6%; Score 18; DB 13; Length 574;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 tcttcattattccaca 118  
|||||  
Db 408 tcttcattattccaca 391

## RESULT 21

AZ243191 595 bp DNA GSS 15-JUN-2000  
DEFINITION RPCI-23-34A15.TV RPCI-23 Mus musculus genomic clone RPCI-23-34A15,  
DNA sequence.

ACCESSION AZ243191  
VERSION AZ243191.1 GI:8556298  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 595)  
AUTHORS Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akimret, B., Levins, M., Mcgann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
COMMENT Other GSS: RPCI-23-34A15.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 34 Row: A Column: 15  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..595  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-34A15"  
/clone.lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). \*"

BASE COUNT 256 a 125 c 74 g 140 t

Query Match 14.6%; Score 18; DB 13; Length 595;  
Best Local Similarity 100.0%; Pred. No. 24;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 101 tcttcattattccaca 118  
|||||  
Db 12 tcttcattattccaca 29

## RESULT 22

BF815581/c 599 bp mRNA EST 13-JAN-2001  
DEFINITION NR2-C10128-041200-009-c02 C10128 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF815581  
VERSION BF815581.1 GI:12150164  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 599)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunsstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=NR2-C10128-041200-009-c02&t3=2000-12-04&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 599.

FEATURES  
source Location/Qualifiers  
1..599  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone.lib="C10128"  
/dev\_stage="Adult"

/note="Organ: colon\_lns; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 156 a 137 c 151 g 152 t 3 others

Query Match 14.6%; Score 18; DB 11; Length 599;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 aattattccacacaca 123  
|||||  
Db 71 aattattccacacaca 54

## RESULT 23

AZ396970 606 bp DNA GSS 03-OCT-2000  
LOCUS AZ396970  
DEFINITION IM016ID13R Mouse 10kb plasmid uucgm library Mus musculus genomic



clone UGCG1M0161D13 R, DNA sequence.  
 ACCESSION A2396970  
 VERSION A2396970.1 GI:10512042  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 REFERENCE 1 (bases 1 to 606)  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A. and Wright,D., Weiss,R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0161 row: D column: 13  
 Seq primer: CACACAGCAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 606.  
 Location/Qualifiers  
 1. 606  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UGCG1M0161D13"  
 /clone\_lib="Mouse 10kb plasmid library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The sheared DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114|g14732114|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 147 a 95 c 107 g 257 t  
 ORIGIN  
 Query Match 14.6%; Score 18; DB 13; Length 606;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ttatctcaattatctcc 115  
 ||||||||||||||||  
 Db 315 TTATCTCAATTTATTC 332

RESULT 24  
 BF208765/c 634 bp mRNA EST 06-NOV-2000  
 LOCUS

601872235f1 NIH\_MGC\_53 Homo sapiens cDNA clone IMAGE:4093065 5', mRNA sequence.  
 ACCESSION BF208765  
 VERSION BF208765.1 GI:11102351  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 REFERENCE 1 (bases 1 to 634)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: c9apbs@remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov  
 Plate: L1CM951 row: f column: 10  
 High quality sequence stop: 489.  
 Location/Qualifiers  
 1. 634  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4093065"  
 /clone\_lib="NIH\_MGC\_53"  
 /tissue\_type="carcinoma, cell line"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: bladder; Vector: pDR-LIB (Clontech); Site\_1: Sili (ggccgcctggcc); Site\_2: Sili (ggccatagcc ); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCCGCGCCGACATG-dT(30)-BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 184 a 176 c 165 g 109 t  
 ORIGIN  
 Query Match 14.6%; Score 18; DB 11; Length 634;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 atgtaagtgatactat 101  
 ||||||||||||||||  
 Db 364 ATCTAGTGATGATCTAT 347

RESULT 25  
 AV326766 167 bp mRNA EST 11-NOV-1999  
 LOCUS AV326766  
 DEFINITION AV326766 RIKEN full-length enriched, adult male medulla oblongata Mus musculus cDNA clone 6330415D21 3', mRNA sequence.  
 ACCESSION AV326766  
 VERSION AV326766.1 GI:6366818  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 REFERENCE 1 (bases 1 to 167)  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,



Seq	Seq	Seq
Oy	28	caaccctgatgaccta 44
Db	25	CAACCTGATGACCTAA 41

VERSION	AV337137.1	GI:63771899
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	







BASE COUNT	87 a	88 c	39 g	63 t
ORIGIN				

Query Match	13.8%;	Score 17;	DB 10;	Length 277;
Best Local Similarity	100.0%;	Pred. No. 77;		
Matches 17; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	28	caaccctgatgacctaa	44
Db	135	CAACCCTGATGACCTAA	151

RESULT	29
AV277980	
LOCUS	AV277980 296 bp mRNA
DEFINITION	AV277980 RIKEN full-length enriched, adult male testis (DH5a) Musculus cdna clone 4932703E19 3', mRNA sequence.

VERSION AV277980.1 GI:6266017  
KEYWORDS EST.

SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
AUTHORS  
Konno, H., Aizawa, K., Akahita, S., Akiyama, J., Carninci, P., Endo, T.,  
1 (bases 1 to 296)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus

Phkuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kato, C., Kawaji, T., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakanuma, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomihata, N., Tsunoda, Y., Watanuki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. (1999) **THE RIKEN MOUSE ESTS (KONNO, H., et al. 1999)**

TITLE	RIKEN MOUSE ESTS (KONNO, H., ET AL. 1999)
JOURNAL	UNPUBLISHED (1999)
COMMENT	CONTACT: YOSHINORI HAYASHIZAKI

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)

Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Iwawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

High-efficiency full-length cDNA cloning. Methods Enzymol. 303: 391-401.

19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

```
location/qualifiers
1..296
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4932703E19"
/clone_lib="RIKEN full-length enriched, adult male testis
(DH5a)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH5a"
```

/note=“Site\_1: Sall; Site\_2 BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5’ GAGAGAGAAGACATCCAGAGCTCTTTTTTTTTTTTTNN 3’], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5’ GAGAGAGAGATTGTCAGGTAAATTAATTAATCCCCCCCCCCC 3’]. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pluviuscript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5’ end: Sall; 3’ end:

BASE COUNT	90 a	91 c	31 g	84 t
ORIGIN				

Query Match	13.84;	Score 17;	DB 10;	Length 296;
Best Local Similarity	100.0%;	Pred. No. 78;		
Matches 17: Conservative	0:	Mismatches	0:	Indels 0:
		Gaps	0:	

QY	28	caaccctgatgacctaa	44
Db	154	CAACCCTGATGACCTAA	170

RESULT	30				
LOCUS	BB248570				
DEFINITION	BB248570	296 bp	mRNA	EST	06-JUL-2000
	BB248570	RIKEN full-length enriched, 7 days neonate cerebellum Mus			
	musculus cDNA clone A730026M11 3', mRNA sequence.				

musculus cDNA clone A7
BB248570
BB248570.1 GI:8941316

**KEYWORDS**  
**SOURCE**  
**ORGANISM**

REFERENCE  
1 (pages 1 to 236)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus

**AUTHORS**  
Kono, H., Alawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carrinici, P., Endo, T., Fukuda, S., Fukunishi, Y., Harai, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadoita, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.

Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Kono, H., et al.)  
Unpublished (2000)



COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki  
, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki  
, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.riken.go.jp>) for  
further details.

FEATURES  
source  
Location/Qualifiers  
1..296  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A730026M11"  
/clone\_lib="RIKEN full-length enriched, 7 days neonate  
cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="7 days neonate"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI. cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGATCCAGACGCTCTTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 20.0 and subtraction to Rot = 459.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence 15' GAGAGAGATCCAGATTAATTAATTAATCCCCCCCCCCC  
3'. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I."

BASE COUNT 97 a 69 c 42 g 88 t  
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 296;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 28 caaccctgatgacctaa 44  
|||||  
Db 156 CAACCTGATGACCTAA 172

RESULT 31  
AI466360 301 bp mRNA EST 09-MAR-1999  
LOCUS  
DEFINITION v888b10.y1 StrataGene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:1262011 5', mRNA sequence.  
ACCESSION AI466360  
VERSION AI466360.1 GI:4314979  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
1 (bases 1 to 301)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, R., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person  
, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter  
, E., Korn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
CONTACT: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
MGI:664563

FEATURES  
source  
Location/Qualifiers  
1..301  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1262011"  
/clone\_lib="Stratagene mouse skin (#937313)"  
/sex="females"  
/tissue\_type="whole skin"  
/dev\_stage="11 weeks old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: skin. Vector: pBluescript SK-; Site 1: EcoRI  
; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
adaptor sequence: 5' GAATTCGACGACGAC 3' -3' adaptor  
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'."

BASE COUNT 62 a 41 c 66 g 130 t  
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 301;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 62 tgcactgactggtgaagt 78  
|||||  
Db 71 TGCACGTGACGTGTAAGT 87

RESULT 32  
BI188737 316 bp mRNA EST 10-JUL-2001  
LOCUS  
DEFINITION d2904fs.r1 Fusarium sporotrichioides Tr1 10 overexpressed cDNA  
library Fusarium sporotrichioides cDNA clone d2904fs 5', mRNA  
sequence.  
ACCESSION BI188737  
VERSION BI188737.1 GI:14662416  
KEYWORDS EST.  
SOURCE Fusarium sporotrichioides.  
ORGANISM Fusarium sporotrichioides.  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocerales; mitosporic Hypocerales; Fusarium.  
REFERENCE  
AUTHORS Ren, Q., Tag, A., Peplow, A., Lal, H., Kupfer, C., Peterson, A., Beremand  
, M. and Roe, B.  
Analysis of a Fusarium sporotrichioides EST database  
Unpublished (2001)



COMMENT  
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
Department of Chemistry and Biochemistry  
Advanced Center for Genome Technology, University of Oklahoma  
620 Parrington Oval, Norman, OK 73019, USA  
Tel: 405 325 4912  
Fax: 405 325 7762  
Email: broe@ou.edu  
Contact Dr. Marian Beremand regarding clone availability included  
is the best homolog from a blastx search of Genbank nr 04-09-01  
[Arabidopsis thaliana]  
Seq primer: T3  
High quality sequence stop: 136.  
Location/Qualifiers  
1. 316  
/organism="Fusarium sporotrichioides"  
/strain="Tr1 10"  
/db\_xref="taxon:5514"  
/clone="q2g04fs"  
/clone\_lib="Fusarium sporotrichioides Tr1 10 overexpressed  
cDNA library"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript  
; 3' end of cDNA cloned into XhoI site of pBluescript"  
BASE COUNT 82 a 102 c 63 g 69 t  
ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 316;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 tgcgcagcgtcactga 69  
|||||  
Db 64 TCGCAGCGCTGCACTGA 48

RESULT 33  
BI266425 329 bp mRNA EST 18-JUL-2001  
LOCUS  
DEFINITION NF097C04IN.F1034 Insect herbivory Medicago truncatula cDNA clone  
ACCESSION BI266425  
VERSION BI266425.1 GI:14870518  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;  
Medicago.  
1 (bases 1 to 329)  
Korth, R., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores  
H.R., Inman, J.T., Weller, J.W. and May, G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula insect herbivory library  
unpublished (2000)  
JOURNAL  
COMMENT  
Contact: Korth R  
Dept. of Plant Pathology  
University of Arkansas  
217 Plant Science Building, Fayetteville, AR 72701, USA  
Tel: 501 575 5191  
Fax: 501 575 7601  
Email: korth@comp.uark.edu  
Insert Length: 329 Std Error: 0.00  
Plate: 097 row: C column: 04  
Seq primer: TCACACAGGAACACGTATGAC.  
FEATURES  
source  
1. 329  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF097C04IN"  
/clone\_lib="insect herbivory"

/tissue\_type="local and systemic leaves"  
/dev\_stage="mature"  
/note="Vector: Lambda Zap. Library was produced from fully  
expanded M. truncatula leaves of plants fed upon by  
Spodoptera exigua (beet armyworm) for 24 hours. Systemic  
(undamaged leaves from injured plants) and wounded leaves  
were harvested and pooled."  
BASE COUNT 111 a 62 c 50 g 104 t 2 others  
ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 329;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 tctcaattatccac 117  
|||||  
Db 42 TCTTCATTATTCAC 58

RESULT 34  
AA244200/c 334 bp mRNA EST 20-AUG-1997  
LOCUS  
DEFINITION nc06c06.s1 NCI-CGAP\_Prl Homo sapiens cDNA clone IMAGE:1007338  
similar to contains Alu repetitive element; contains element MER35  
repetitive element ;, mRNA sequence.  
ACCESSION AA244200  
VERSION AA244200.1 GI:1874959  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 334)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index.  
Unpublished (1997)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@emil.nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui, M.D.  
Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: -41ml3 fwd. ET from Amersham  
High quality sequence stop: 293.  
Location/Qualifiers  
1. 334  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1007338"  
/clone\_lib="NCI-CGAP\_Prl"  
/sex="male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from 5,000-10  
000 microdissected, histologically normal prostate  
epithelial cells. Double-stranded cDNA was ligated to  
EcoRI adaptors, 5 cycles of PCR applied to the cDNA with  
an adaptor-specific primer, and the resulting PCR product  
subcloned into PAMP10 by the upc-cloning method (Life  
technologies). Average insert size is 600 bp. NOTE: Not  
directionally cloned. This library was constructed by  
David Krizman."

BASE COUNT 111 a 66 c 82 g 75 t  
ORIGIN



Query Match 13.8%; Score 17; DB 10; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 99 tatcttaattattcc 115  
 |||  
 Db 33 TATCTCAATTATTC 17

RESULT 35  
 C70635/c 360 bp mRNA EST 29-SEP-1997  
 LOCUS C70635 yuji Kohara unpublished cdna Caenorhabditis elegans cDNA  
 DEFINITION clone yK411b1 5', mRNA sequence.  
 ACCESSION C70635  
 VERSION C70635.1 GI:2441160  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea  
 ; Rhabditidae; Pelodderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 360)  
 AUTHORS Kohara,Y., Morohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano  
 ,M., Miyata,A. and Nishigaki,A.  
 TITLE Expression map of the C.elegans genome  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Yuji Kohara  
 Genome Biology Lab.  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 81-559-81-6854  
 Fax: 81-559-81-6855  
 Email: ykohara@ab.nig.ac.jp.

FEATURES  
 source Location/Qualifiers  
 1..360  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="YK411b1"  
 /clone.lib="Yuji Kohara unpublished cdna"  
 /sex="hermaphrodite, male"  
 /tissue\_type="whole animal"  
 /dev\_stage="varied"

BASE COUNT 102 a 63 c 79 g 106 t 10 others

ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 105 caattattcacacaa 121  
 |||  
 Db 287 CAATTATTCACACAA 271

RESULT 36  
 B1266362 366 bp mRNA EST 18-JUL-2001  
 LOCUS B1266362  
 DEFINITION NF092D091NF1078 Insect herbivory Medicago truncatula cDNA clone  
 NF092D091N 5', mRNA sequence.  
 ACCESSION B1266362  
 VERSION B1266362.1 GI:14870403  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.  
 REFERENCE 1 (bases 1 to 366)  
 AUTHORS Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores

TITLE ,H.R., Inman,J.T., Weller,J.W. and May,G.D.  
 JOURNAL Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
 COMMENT Medicago truncatula insect herbivory library  
 Unpublished (2000)  
 Contact: Korth K  
 Dept. of Plant Pathology  
 University of Arkansas  
 217 Plant Science Building, Fayetteville, AR 72701, USA  
 Tel: 501 575 5191  
 Fax: 501 575 7601  
 Email: Kkorth@comp.uark.edu  
 Insert Length: 366 Std Error: 0.00  
 Plate: 092 row: D column: 09  
 Seq primer: TCACACGACGAACACCTATGAC.  
 Location/Qualifiers  
 1..366  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone="NF092D091N"  
 /clone.lib="Insect herbivory"  
 /tissue\_type="local and systemic leaves"  
 /dev\_stage="mature"  
 /note="Vector: Lambda Zap; Library was produced from fully  
 expanded M. truncatula leaves of plants fed upon by  
 Spodoptera exigua (beet armyworm) for 24 hours. Systemic  
 (undamaged leaves from injured plants) and wounded leaves  
 were harvested and pooled."

BASE COUNT 115 a 69 c 52 g 128 t 2 others

ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 101 tcttcaattattccac 117  
 |||  
 Db 104 TCTTCAATTATTCAC 120

RESULT 37  
 AM457639 450 bp mRNA EST 24-FEB-2000  
 LOCUS AM457639  
 DEFINITION UI-M-BH3-ari-a-04-0-01.s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
 UI-M-BH3-ari-a-04-0-01 3', mRNA sequence.  
 ACCESSION AM457639  
 VERSION AM457639.1 GI:7027896  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 450)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Chiu, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mestr@nim.nih.gov  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized corpus striatum library cDNA Library Preparation: M.B.  
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares



is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLY-A-Tes.

#### FEATURES

source

Location/Qualifiers  
1. 456  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-ari-a-04-0-UI"  
/clone\_lib="NIH\_BMAP\_M\_S4"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI. The NIH\_BMAP\_M\_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH\_BMAP\_M\_S4, NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1, NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH\_BMAP\_M\_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_TISSUE=corpus-striatum  
TAG\_SEQ=ACGCGC"

BASE COUNT 110 a 127 c 109 g 104 t  
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 450;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 gactggaatttcaca 84  
|||||  
DB 377 GACTGGTAGTTTCACA 393

#### RESULT 38

LOCUS BE649642 456 bp mRNA EST 20-DEC-2000  
DEFINITION NF082D01EC1F1012 Elicited cell culture Medicago truncatula cDNA  
ACCESSION BE649642  
VERSION BE649642.1 GI:11914772  
KEYWORDS EST  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.  
1 (bases 1 to 456)  
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

TITLE  
JOURNAL  
COMMENT  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research  
Unpublished (2000)  
Contact: Dixon RA  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7302  
Fax: 580 221 7380  
Email: radixon@noble.org  
Insert Length: 456 Std Error: 0.00  
Plate: 082 row: D column: 01  
Seq primer: TCACACGAGAACACGCTATGAC.

#### FEATURES

source

Location/Qualifiers  
1. 456  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF082D01EC"  
/clone\_lib="Elicited cell culture"  
/tissue\_type="Cell cultures derived from root tissues"  
/dev\_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"  
/note="Vector: Lambda Zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5, 1, 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."

BASE COUNT 146 a 86 c 66 g 158 t  
ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 456;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 tcttcaattatccac 117  
|||||  
DB 120 TCTTCATTATTCAC 136

#### RESULT 39

LOCUS BE989827 489 bp mRNA EST 05-OCT-2000  
DEFINITION UI-M-BZ1-bft-e-20-0-UI s1 NIH\_BMAP\_MHT12\_S1 Mus musculus cDNA clone  
ACCESSION BE989827  
VERSION BE989827.1 GI:10667610  
KEYWORDS EST  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 489)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
JOURNAL MEDLINE  
COMMENT  
97044477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mestr@mail.nih.gov  
The sequence contained an oligo-dr track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to verify it as a clone from the hippocampus tissue cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a



small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLY-A=yes

FEATURES  
source

Location/Qualifiers  
1. 489  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BZ1-Dft-e-20-0-UI"  
/clone\_11b="NIH\_BMAP\_MHI2\_S1"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The NIH\_BMAP\_MHI2\_S1 library is a subtracted library derived from NIH\_BMAP\_MHI2. NIH\_BMAP\_MHI2 is a library derived from mouse hippocampus tissue. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.  
TAG\_LIB=NIH\_BMAP\_MHI2\_S1  
TAG\_TISSUE=hippocampus  
TAG\_SEQ=TACTC"

BASE COUNT 118 a 142 c 123 g 106 t  
ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 489;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 gactgtaagttcaca 84  
|||||  
Db 356 GACTGTAAGTTTACA 372

RESULT 40  
AM394535 519 bp mRNA EST 07-FEB-2000  
LOCUS sh32c05.y1 Gm-cl017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
DEFINITION Gm-cl017-3393 5' similar to TR:083191 083191 CONSERVED HYPOTHETICAL  
PROTEIN ; mRNA sequence.  
ACCESSION AM394535  
VERSION AM394535.1 GI:6912947  
KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.  
1 (bases 1 to 519)  
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna  
A., Bolla,B., Mair,M., Hiller,L., Kucaba,T., Martin,J., Beck,C.,  
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
,R., Waterston,R. and Wilson,R.  
Public Soybean EST project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST project  
Public Soybean EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

TITLE JOURNAL  
COMMENT This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: www.genomesystems.com

Seq primer: -40RP from Gibco  
High quality sequence stop: 426.  
Location/Qualifiers  
1. 519  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-3393"  
/clone\_11b="Gm-cl017"  
/tissue\_type="vegetable buds of field grown plants"  
/lab\_host="Xl10-Gold"  
/note="Vector: p Bluescript II Xr; Site\_1: EcoRI; Site\_2:  
XhoI. This cDNA library was constructed from mRNA isolated  
from vegetable buds of field grown plants. The cDNA  
library was prepared using the Stratagene p Bluescript II  
Xr library construction kit. Complementary DNA was  
synthesized from mRNA using a primer consisting of a poly  
(dT) sequence with a XhoI restriction site. EcoRI adapters  
were ligated to the blunt-ended cDNA fragments followed by  
XhoI digestion. The cDNA fragments were directionally  
cloned into the EcoRI-XhoI restriction site of the  
p Bluescript vector. The ligated cDNA fragments were  
transformed into Xl10-Gold host cells. This library was  
constructed by Dr. Randy Shoemaker and Dr. John  
Erpelting."

BASE COUNT 143 a 91 c 120 g 164 t 1 others  
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 519;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 tactatctcaattta 111  
|||||  
Db 488 TACTATCTCAATTTA 504

RESULT 41  
A2626769 526 bp DNA GSS 13-DEC-2000  
LOCUS A2626769/c  
DEFINITION 1M0467008F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0467008 F, DNA sequence.  
ACCESSION A2626769  
VERSION A2626769.1 GI:11748959  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 526)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0467 row: 0 column: 08  
Seq primer: CGTTGTAAACGACGCGCAGT  
Class: plasmid ends  
High quality sequence stop: 526.  
Location/Qualifiers  
1. 526  
/organism="Mus musculus"

FEATURES  
source



```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="UGC1M0467008"
/clone_1lb="Mouse 10kb plasmid UGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      215 a      97 c      92 g      122 t
ORIGIN

```

```

Query Match      13.8%; Score 17; DB 13; Length 526;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 100 attccaattatcca 116
      |||||
Db 179 ATCTCAATTATTCCA 163

```

```

RESULT 42
BG048538      541 bp      mRNA      EST      25-JAN-2001
LOCUS
DEFINITION
OVL_14_H02_g2_A002 Ovary 1 (OVL) Sorghum bicolor cDNA, mRNA
sequence:
ACCESSION
BG048538.1
VERSION
BG048538.1
KEYWORDS
EST.
SOURCE
sorghum.
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 541)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
An EST database from Sorghum: ovaries of varying immature stages
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@prattuga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTmix
High quality sequence start: 30
High quality sequence stop: 541
POLYA-NO.
Location/Qualifiers
1. 541
/organism="Sorghum bicolor"

```

```

FEATURES
source
1. 541
/organism="Sorghum bicolor"

```

```

/db_xref="taxon:4558"
/clone_1lb="Ovary 1 (OVL)"
/notes="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT      139 a      111 c      141 g      150 t
ORIGIN

```

```

Query Match      13.8%; Score 17; DB 11; Length 541;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 26 tacacacctgagct 42
      |||||
Db 74 TACACACCTGATGACT 90

```

```

RESULT 43
AZ961481      544 bp      DNA      GSS      27-APR-2001
LOCUS
DEFINITION
2M0229L22R Mouse 10kb plasmid UGC2M library Mus musculus genomic
clone UGC2M0229L22 R; DNA sequence.
ACCESSION
AZ961481
VERSION
AZ961481.1
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS
1 (bases 1 to 544)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingy,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0229 row: 1 column: 22
Seq primer: CACACAGGAACACGCTATGACC
Class: Plasmid ends
High quality sequence stop: 544.
Location/Qualifiers
1. 544
/organism="Mus musculus"

```

```

FEATURES
source
1. 544
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="UGC2M0229L22"
/clone_1lb="Mouse 10kb plasmid UGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel

```



electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473214|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 166 a 139 c 101 g 138 t  
ORIGIN

Query Match 13.8%; Score 17; DB 13; Length 544;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 69 actgtaagtccat 85  
|||||  
Db 426 ACTGTAAGTTTCACAT 442

RESULT 44  
A2561989

LOCUS 555 bp DNA GSS 20-NOV-2000  
DEFINITION RPCI-23-21704.TV RPCI-23 Mus musculus genomic clone RPCI-23-21704,  
DNA sequence.

ACCESSION A2561989

VERSION A2561989.1 GI:11241809

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 555)

REFERENCE Zhao,S., Niemman,W., Feldblum,T., Malek,J., Shatsman,S., Akimret,  
B., Levins,M., McGann,S., Tsengaye,G., Geer,K., Krol,M., de Jong,P.  
and Fraser,C.M.

Mouse BAC End Sequences from library RPCI-23  
Unpublished (1999)

Other GSSs: RPCI-23-21704.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.bufileo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)

or from ResGen Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tldb/bac-ends/mouse/bac\_end\_intro.html

Plate: 217 row: 0 column: 4

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..555

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-21704"

/clone\_1lb="RPCI-23"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pAC3.6; Site:1;

EcoRI; Site:2; EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pAC3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT

193 a 108 c 116 g 138 t

ORIGIN

Query Match 13.8%; Score 17; DB 13; Length 555;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 tgcactgactgtaagt 78  
|||||  
Db 535 TGCACGTACTGTGTAAGT 551

RESULT 45

A2149056

LOCUS 557 bp DNA GSS 28-AUG-2000

DEFINITION SP.0046.AL.B09.T7 Strongylocentrotus purpuratus, purple sea urchin,  
sperm genomic BAC library Strongylocentrotus purpuratus genomic

clone plate=46 Col=17 Row=C, DNA sequence.

ACCESSION A2149056

VERSION A2149056.1 GI:8300957

KEYWORDS GSS.

SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

Echinoidea; Euechinoidea; Echinodermata; Echinodermata;

Strongylocentrotidae; Strongylocentrotus.

1 (bases 1 to 557)

REFERENCE Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,  
Swartzell,S., Wallace,J.C., Pousetka,A.J., Livingston,B.T., Wray,  
G.A., Eitensohn,C.A., Lehrach,H., Bittlen,R.J., Davidson,E.H. and  
Hood,L.

A sea urchin genome project: Sequence scan, virtual map, and  
additional resources

Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

20402566

Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 46 row: C column: 17

Seq primer: T7

Class: BAC ends

High quality sequence shop: 557.

Location/Qualifiers

1..557

/organism="Strongylocentrotus purpuratus"

/db\_xref="taxon:7668"

/clone="Plate=46 Col=17 Row=C"

/clone\_1lb="Strongylocentrotus purpuratus, purple sea

urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli

DH10B"

BASE COUNT 177 a 111 c 114 g 152 t 3 others

ORIGIN

Query Match 13.8%; Score 17; DB 13; Length 557;

Best Local Similarity 100.0%; Pred. No. 82;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 tatttgaatgtagat 21

|||||

Db 455 TATTGTAATGTGACAT 471

Search completed: March 19, 2002, 10:40:41

Job time: 3945 sec







**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:58:46 ; Search time 1344.34 Seconds

(without alignments)  
1509.405 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 caggtatgtgtaaatgtgaga.....tcatttattccacacacaca 123

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgt\_hum:\*  
31: em\_hgtgo\_hum:\*  
32: em\_hgtgo\_inv:\*  
33: em\_hgtgo\_rod:\*  
34: em\_hgtgo\_hum:\*  
35: em\_hgtgo\_inv:\*  
36: em\_hgtgo\_rod:\*  
37: em\_hgtgo\_inv:\*  
38: em\_hgtgo\_rod:\*  
39: em\_hgtgo\_hum:\*  
40: em\_hgtgo\_inv:\*  
41: em\_hgtgo\_rod:\*  
42: em\_hgtgo\_hum:\*  
43: em\_hgtgo\_inv:\*  
44: em\_hgtgo\_rod:\*  
45: em\_hgtgo\_hum:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	17.1	69734	2	AC010953	AC010953 Homo sapi
2	21	17.1	178035	2	AC068680	AC068680 Homo sapi
3	21	17.1	183079	2	AC007600	AC007600 Homo sapi
4	21	17.1	194198	2	AC013298	AC013298 Homo sapi
5	20	16.3	50974	9	AP000958	AP000958 Homo sapi
6	20	16.3	155304	2	AC024713	AC024713 Homo sapi
7	20	16.3	158927	2	CNS07EFA	CNS07EFA Homo sapi
8	20	16.3	171345	9	CNS01RIT	CNS01RIT Homo sapi
9	20	16.3	182340	2	AC024417	AC024417 Homo sapi
10	20	16.3	340000	9	AP001675	AP001675 Homo sapi
11	19	15.4	27710	3	U41746	U41746 Caenorhabd
12	19	15.4	130057	9	AL357912	AL357912 Human DNA
13	19	15.4	180911	2	AC027498	AC027498 Homo sapi
14	18	14.6	1103	6	AX073208	AX073208 Sequence
15	18	14.6	11018	1	AE002395	AE002395 Neisseria
16	18	14.6	33803	3	CEL34H3	CEL34H3 Caenorhab
17	18	14.6	34128	9	AC022128	AC022128 Homo sapi
18	18	14.6	57000	9	AC078862	AC078862 Homo sapi
19	18	14.6	80374	8	T8K14	T8K14 Homo sapi
20	18	14.6	117018	9	AC008790	AC008790 Homo sapi
21	18	14.6	140952	8	AP003046	AP003046 Oryza sat
22	18	14.6	144536	9	AL161454	AL161454 Human DNA
23	18	14.6	146571	2	AL359815	AL359815 Homo sapi
24	18	14.6	148498	2	AC034225	AC034225 Homo sapi
25	18	14.6	153405	2	AL357252	AL357252 Homo sapi
26	18	14.6	156795	2	AC023537	AC023537 Homo sapi
27	18	14.6	165858	2	AC024315	AC024315 Homo sapi
28	18	14.6	183496	2	AC009776	AC009776 Homo sapi
29	18	14.6	192430	2	AC068581	AC068581 Homo sapi
30	18	14.6	192430	2	AC079542	AC079542 Mus muscu
31	18	14.6	198605	2	AL357568	AL357568 Homo sapi
32	18	14.6	218073	2	AC025501	AC025501 Mus muscu
33	18	14.6	226060	2	AC079583	AC079583 Mus muscu
34	18	14.6	349980	6	AX044029	AX044029 Sequence
35	17	13.8	518	6	AX094489	AX094489 Sequence
36	17	13.8	2195	20	AY004261	AY004261 Heterocap
37	17	13.8	34544	3	CELY5868A	CELY5868A Caenorhab
38	17	13.8	38225	9	AL591603	AL591603 Human DNA
39	17	13.8	44751	8	SPC116C4	SPC116C4 S.pombe c
40	17	13.8	50821	8	AC020579	AC020579 Arabidops
41	17	13.8	59935	2	AC018376	AC018376 Homo sapi
42	17	13.8	63604	8	AP000381	AP000381 Arabidops
43	17	13.8	65857	2	AC018413	AC018413 Homo sapi
44	17	13.8	65998	2	AC023633	AC023633 Homo sapi
45	17	13.8	70677	2	AC027574	AC027574 Homo sapi

## ALIGNMENTS

RESULT	1.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
--------	----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



TITLE	Direct Submission
JOURNAL	Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. &

Forrester, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardy, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, B., Katarata, A., Klein, J., Lechoczky, J., Liu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGuck, A., McKenken, K., McNaughton, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Tirelli, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

\* NOTE: This record contains 78 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

1	933:	contig of 933 bp in length
*		gap of unknown length
*	934	1798: contig of 865 bp in length
*		gap of unknown length
*	1799	2683: contig of 885 bp in length
*		gap of unknown length
*	2684	3569: contig of 886 bp in length
*		gap of unknown length
*	3570	4443: contig of 874 bp in length
*		gap of unknown length
*	4444	5297: contig of 854 bp in length
*		gap of unknown length
*	5298	6183: contig of 886 bp in length
*		gap of unknown length
*	6184	7057: contig of 874 bp in length
*		gap of unknown length
*	7058	7989: contig of 932 bp in length
*		gap of unknown length
*	7990	8859: contig of 870 bp in length
*		gap of unknown length
*	8860	9721: contig of 862 bp in length
*		gap of unknown length
*	9722	10593: contig of 872 bp in length
*		gap of unknown length
*	10594	11472: contig of 879 bp in length
*		gap of unknown length
*	11473	12381: contig of 909 bp in length
*		gap of unknown length
*	12382	13275: contig of 894 bp in length
*		gap of unknown length
*	13276	14178: contig of 903 bp in length
*		gap of unknown length
*	14179	15072: contig of 894 bp in length
*		gap of unknown length
*	15073	15954: contig of 882 bp in length
*		gap of unknown length
*	15955	16816: contig of 862 bp in length
*		gap of unknown length
*	16817	17750: contig of 934 bp in length
*		gap of unknown length
*	17751	18613: contig of 863 bp in length
*		gap of unknown length
*	18614	19490: contig of 877 bp in length
*		gap of unknown length
*	19491	20372: contig of 882 bp in length
*		gap of unknown length

20373	21256	contig of 886 bp in length
21259	22199	contig of 941 bp in length
22200	23107	gap of unknown length
23108	24014	contig of 908 bp in length
24015	24893	gap of unknown length
24894	25753	contig of 879 bp in length
25754	26631	gap of unknown length
26632	27505	contig of 878 bp in length
27506	28396	gap of 874 bp in length
28397	29283	gap of unknown length
29284	30172	contig of 891 bp in length
30173	31120	gap of unknown length
31121	32109	contig of 948 bp in length
32110	33003	gap of unknown length
33004	33886	contig of 884 bp in length
33887	34768	gap of 883 bp in length
34769	35647	gap of unknown length
35648	36542	contig of 882 bp in length
36543	37411	gap of unknown length
37412	38287	contig of 879 bp in length
38288	39172	gap of 876 bp in length
39173	40064	gap of unknown length
40065	40944	contig of 885 bp in length
40945	41799	gap of unknown length
41800	42711	contig of 855 bp in length
42712	43589	gap of unknown length
43590	44511	contig of 912 bp in length
44512	45377	gap of 878 bp in length
45378	46252	gap of unknown length
46253	47126	contig of 866 bp in length
47127	48022	gap of unknown length
48023	48921	contig of 874 bp in length
48922	49794	gap of 899 bp in length
49795	50651	contig of 873 bp in length
50652	51530	gap of unknown length
51531	52415	contig of 857 bp in length
52416	53304	gap of 879 bp in length
53304		gap of unknown length
		contig of 885 bp in length
		gap of 889 bp in length
		contig of 889 bp in length







```
misc_feature 1..1212
              /note="assembly_fragment"
              clone_end:Sp6
              vector_side:right"
misc_feature 1313..5818
              /note="assembly_fragment"
              5919..9462
              /note="assembly_fragment"
              9563..15211
              /note="assembly_fragment"
              15312..28255
              /note="assembly_fragment"
              28356..44646
              /note="assembly_fragment"
              44747..63409
              /note="assembly_fragment"
              63510..90494
              /note="assembly_fragment"
              clone_end:T7
misc_feature 90595..118529
              /note="assembly_fragment"
              118630..145844
              /note="assembly_fragment"
              145945..178035
              /note="assembly_fragment"
BASE COUNT 48869 a 40236 c 39973 g 47951 t 1006 others
ORIGIN
```

```
Query Match 17.1%; Score 21; DB 2; Length 178035;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 98 ttatctcaattatccaca 118
|||||
Db 169655 TTATCTCAATTATTCACA 169655
```

```
RESULT 3
AC007600/c AC007600 183079 bp DNA HTG 03-JUL-2001
LOCUS Homo sapiens chromosome 16 clone RP11-264A16, WORKING DRAFT
DEFINITION
AC007600
AC007600.4 GI:14589421
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE 1 (bases 1 to 183079)
JOURNAL DOE Joint Genome Institute.
REFERENCE Sequencing of Human Chromosome 16
AUTHORS Unpublished
TITLE 2 (bases 1 to 183079)
JOURNAL Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
REFERENCE Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
AUTHORS Bryant, J., Tesmer, J., Meinke, L., Longmire, D., White, S., Tatum, O.,
JOURNAL Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,
REFERENCE McMurry, K., Han, C. and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
REFERENCE Genome Institute, Los Alamos National Laboratory, MS M888, Los
AUTHORS Alamos, NM 87545, USA
COMMENT On Jul 3, 2001 this sequence version replaced gi:8575868.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
```

GenBank flat file format but are available as part of this entry's ASN.1 file.

#### Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 183079: contig of 183079 bp in length.

```
FEATURES
    source
        1..183079
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /clone="RP11-264A16"
BASE COUNT 47478 a 41752 c 42357 g 51492 t
ORIGIN
```

```
Query Match 17.1%; Score 21; DB 2; Length 183079;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 98 ttatctcaattatccaca 118
|||||
Db 130958 TTATCTCAATTATTCACA 130958
```

```
RESULT 4
AC013298 AC013298 194198 bp DNA HTG 20-SEP-2000
LOCUS Homo sapiens clone RP11-3N3, WORKING DRAFT SEQUENCE, 19 unordered
DEFINITION
AC013298
AC013298.6 GI:10198404
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE 1 (bases 1 to 194198)
JOURNAL Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE 2 (bases 1 to 194198)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Baldwin, J., Barna, N., Beckler, R., Boguslavsky, L., Boukhalter, B.,
REFERENCE Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
AUTHORS Cooke, P., Deatellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
JOURNAL Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
REFERENCE Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
AUTHORS Howland, J. C., Johnson, R., Jones, C., Kamp, L., Karas, A., Klein, J.,
JOURNAL Lehotzky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
REFERENCE McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
AUTHORS Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
JOURNAL Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
REFERENCE Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
AUTHORS Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
JOURNAL Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
```



TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 20, 2000 this sequence version replaced gi:9121366.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: L2788

Center clone name: 3.N.3

## Summary Statistics

Sequencing vector: MJ3; M77815, 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 185739 bases at least Q40  
Consensus quality: 189725 bases at least Q30  
Consensus quality: 191400 bases at least Q20  
Insert size: 19400; agarose-fp  
Insert size: 192398; sum-of-contigs  
Quality coverage: 5.4 in Q20 bases; agarose-fp  
Quality coverage: 5.5 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 16747: contig of 16747 bp in length
16748 16847: gap of 100 bp
16848 18184: contig of 1337 bp in length
18185 18284: gap of 100 bp
18285 21234: contig of 2950 bp in length
21235 21334: gap of 100 bp
21335 23492: contig of 2158 bp in length
23493 23592: gap of 100 bp
23593 26852: contig of 3260 bp in length
26853 26952: gap of 100 bp
26953 30877: contig of 3925 bp in length
30878 30977: gap of 100 bp
30978 35870: contig of 4893 bp in length
35871 35970: gap of 100 bp
35971 40691: contig of 4721 bp in length
40692 40791: gap of 100 bp
40792 44904: contig of 4113 bp in length
44905 45004: gap of 100 bp
45005 51760: contig of 6756 bp in length
51761 51860: gap of 100 bp
51861 58723: contig of 6863 bp in length
58724 58823: gap of 100 bp
58824 68160: contig of 9337 bp in length
68161 68260: gap of 100 bp
68261 84475: contig of 16215 bp in length
84476 84575: gap of 100 bp
84576 98534: contig of 13959 bp in length
98535 98634: gap of 100 bp
98635 114643: contig of 16009 bp in length
114644 114743: gap of 100 bp
114744 144676: contig of 29933 bp in length
144677 144776: gap of 100 bp
144777 160047: contig of 15271 bp in length
160048 160147: gap of 100 bp
160148 178836: contig of 18689 bp in length
178837 178935: gap of 100 bp
178937 194198: contig of 15262 bp in length.
Location/Qualifiers

```

```

source
1. 194198
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-383"
/clone_lib="RP11-11 Human Male BAC"
1. 16747
/clone="assembly-fragment"
clone_end:SP6
vector_side:left"
16848. 18184
/clone="assembly-fragment"
18285. 21234
/clone="assembly-fragment"
21335. 23492
/clone="assembly-fragment"
23593. 26852
/clone="assembly-fragment"
26953. 30877
/clone="assembly-fragment"
30978. 35870
/clone="assembly-fragment"
35971. 40691
/clone="assembly-fragment"
40792. 44904
/clone="assembly-fragment"
45005. 51760
/clone="assembly-fragment"
51861. 58723
/clone="assembly-fragment"
58824. 68160
/clone="assembly-fragment"
68261. 84475
/clone="assembly-fragment"
84576. 98534
/clone="assembly-fragment"
98635. 114643
/clone="assembly-fragment"
114744. 144676
/clone="assembly-fragment"
144777. 160047
/clone="assembly-fragment"
160148. 178836
/clone="assembly-fragment"
178937. 194198
/clone="assembly-fragment"
clone_end:77
vector_side:right"
BASE COUNT 52345 a 44636 c 44383 g 51032 t 1802 others
ORIGIN

```

```

Query Match 17.1% Score 21: DB 2: Length 194198;
Best Local Similarity 100.0%, Pred. No. 0.31;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 98 ttatctcaattatccaca 118
|||||
DB 104368 TTATCTCAATTATTCACCA 104388

```

```

RESULT 5
AP000958/c
LOCUS Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B794P4,
DEFINITION Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B794P4,
ACCESSION AP000958
VERSION AP000958.2 GI:6705920
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:B794P4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 50974)

```



**AUTHORS** Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
**TITLE** Homo sapiens 50,974 genomic DNA of 21q21.1-q21.2  
**JOURNAL** Published Only in Database (1999) In press  
**REFERENCE** 2 (bases 1 to 50974)  
**AUTHORS** Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (17-DEC-1999) to the DDBJ/EMBL/GenBank databases.  
**REFERENCE** Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan  
 (E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)  
**COMMENT** On Jan 16, 2000 this sequence version replaced gi:630658.  
 Sequence updated (13-Jan-2000).  
**FEATURES** Location/Qualifiers  
**SOURCE** 1. 50974  
 /Organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /Chromosome="21"  
 /Clone="B794P4"  
 /map="21q21.1-q21.2"  
**BASE COUNT** 16566 a 8657 c 8595 g 17156 t  
**ORIGIN**  
 Query Match 16.3%; Score 20; DB 9; Length 50974;  
 Best Local Similarity 100.0%; Pred.No.1.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 97 ctatctcaattatcca 116  
 ||||||||||||||||  
 Db 21415 CTTATCTCAATTATTCCA 21396  
**RESULT** 6  
**AC024713** AC024713 155304 bp DNA 01-SEP-2000  
**LOCUS** Homo sapiens chromosome 21 clone RP11-170F21, WORKING DRAFT  
**DEFINITION** AC024713  
**ACCESSION** AC024713.4 GI:9958179  
**VERSION** HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
**KEYWORDS** human.  
**SOURCE** Homo sapiens  
**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 155304)  
**AUTHORS** Waterston,R.H.  
**TITLE** The sequence of Homo sapiens clone  
**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 155304)  
**AUTHORS** Waterston,R.H.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (01-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
**COMMENT** On Sep 1, 2000 this sequence version replaced gi:8569835.  
 ----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site:http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H.NH0170F21  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 100%  
 Sequencing vector: plasmid; 0%  
 Chemistry: Dye-Primer ET; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 0% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 138354 bases at least Q40

Consensus quality: 142506 bases at least Q30  
 Consensus quality: 144038 bases at least Q20  
 Insert size: 150000; agarose-fp  
 Insert size: 150578; sum-of-contigs  
 Quality coverage: 3.56 in Q20 bases; agarose-fp  
 Quality coverage: 3.67 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 31 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 900 899: contig of 899 bp in length  
 1000 999: gap of unknown length  
 1827 1826: contig of 827 bp in length  
 1927 1926: gap of unknown length  
 3373 3372: contig of 1447 bp in length  
 3374 3473: gap of unknown length  
 3474 4709: contig of 1236 bp in length  
 4710 4809: gap of unknown length  
 4810 5988: contig of 1179 bp in length  
 5989 6088: gap of unknown length  
 6089 8166: contig of 2078 bp in length  
 8167 8267 8266: gap of unknown length  
 10821 10820: contig of 2554 bp in length  
 10920 10920: gap of unknown length  
 10921 12468: contig of 1548 bp in length  
 12469 12568: gap of unknown length  
 12569 14508: contig of 1940 bp in length  
 14509 14608: gap of unknown length  
 14609 16851: contig of 2243 bp in length  
 16852 16951: gap of unknown length  
 16952 19153: contig of 2202 bp in length  
 19154 19253: gap of unknown length  
 19254 22560: contig of 3307 bp in length  
 22561 22660: gap of unknown length  
 22661 25965: contig of 3305 bp in length  
 25966 26065: gap of unknown length  
 26066 30499: contig of 4434 bp in length  
 30500 30599: gap of unknown length  
 30600 33411: contig of 2812 bp in length  
 33412 33511: gap of unknown length  
 33512 37042: contig of 3531 bp in length  
 37043 37142: gap of unknown length  
 37143 40035: contig of 2893 bp in length  
 40036 40135: gap of unknown length  
 40136 45130: contig of 4995 bp in length  
 45131 45230: gap of unknown length  
 45231 49657: contig of 4427 bp in length  
 49658 49757: gap of unknown length  
 49758 54058: contig of 4301 bp in length  
 54059 54158: gap of unknown length  
 54159 58730: contig of 4572 bp in length  
 58731 58830: gap of unknown length  
 58831 64337: contig of 5507 bp in length  
 64338 64437: gap of unknown length  
 64439 69738: contig of 5301 bp in length  
 69739 69838: gap of unknown length  
 69839 76434: contig of 6596 bp in length  
 76435 76534: gap of unknown length  
 76535 84858: contig of 8324 bp in length  
 84859 92762: gap of unknown length  
 92763 92862: contig of 7804 bp in length  
 92863 92862: gap of unknown length  
 100410 100509: gap of 7547 bp in length  
 100510 109405: gap of unknown length  
 109406 109505: contig of 8896 bp in length  
 109506 121992: gap of unknown length  
 121993 122092: contig of 12487 bp in length  
 122093 122092: gap of unknown length



FEATURES	*	122093	134408:	contig of 12316 bp	in length
	*	134409	134508:	gap of unknown length	
	*	134509	155304:	contig of 20796 bp	in length
	Location/Qualifiers				
	1..155304				
	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /clone="RP11-170F21"				
SOURCE					
misc_feature		1.. 899			
	/note="assembly_name:Contig10"	1000..1826			
	/note="assembly_name:Contig9"	1927..3373			
	/note="assembly_name:Contig16"	3474..4709			
	/note="assembly_name:Contig17"	4810..5988			
	/note="assembly_name:Contig18"	6089..8166			
	/note="assembly_name:Contig19"	8267..10820			
	/note="assembly_name:Contig20"	10921..12468			
	/note="assembly_name:Contig21"	12569..14508			
	/note="assembly_name:Contig22"	14609..16851			
	/note="assembly_name:Contig23"	16952..19153			
	/note="assembly_name:Contig24"	19254..22560			
	/note="assembly_name:Contig25"	22661..25965			
	/note="assembly_name:Contig26"	26066..30499			
	/note="assembly_name:Contig27"	30600..33411			
	/note="assembly_name:Contig28"	33512..37042			
	/note="assembly_name:Contig29"	37143..40035			
	/note="assembly_name:Contig30"	40136..45130			
	/note="assembly_name:Contig31"	45331..49657			
	/note="assembly_name:Contig32"	49758..54058			
	/note="assembly_name:Contig33"	54159..58730			
	/note="assembly_name:Contig34"	58831..64337			
	/note="assembly_name:Contig35"	64438..69738			
	/note="assembly_name:Contig36"	69839..76434			
	/note="assembly_name:Contig37"	76535..84858			
	/note="assembly_name:Contig38"	84959..92762			
	/note="assembly_name:Contig39"	92863..100409			
	/note="assembly_name:Contig40"	100510..109405			
	/note="assembly_name:Contig41"	109506..121992			
	/note="assembly_name:Contig42"	122093..134408			
	/note="assembly_name:Contig43"	134509..155304			
	/note="assembly_name:Contig44"	26013 c 26393 g 49371 t			
BASE COUNT	50522 a	3005 others			
ORIGIN					

	Query Match	Score	DB	Length
Best Local Similarity	100.0%	Pred. No. 1.1	155304	
Matches	20	Conservative	0	Mismatches
			0	Indels
			0	Gaps
QY	97	cttattctcaattatttcca	116	
Db	47722	cttattctcaattatttcca	47741	

RESULT 7	LOCUS CNS07EFA/c	DEFINITION Homo sapiens chromosome 14, clone R-326E7, *** PROGRESS ***, 2 ordered places.	HTNG 29-MAY-2000
ACCESSION	AL512361.2	GI:13548503	
VERSION	HTG; HTGS_PHASE2; HTGS_ACTIVEPIN; HTGS_DRAFT.		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 158927)		
AUTHORS	Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker		

TITLE	Sequencing of the human chromosome 14
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 158927)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (22-MAY-2001) Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	On Apr 4, 2001 this sequence version replaced gi:11990876.

Overall quality chart :		
Range	:	Bases
0 - 9	:	598
1 - 9	:	17
10 - 19	:	203
20 - 29	:	567
30 - 39	:	1808
40 - 49	:	3526
50 - 59	:	7642
60 - 69	:	15666
70 - 79	:	36142
80 - 89	:	53454
90 - 99	:	37304

----- Genome Center  
Center: Genoscope / Centre National de Sequencage  
Center code: GS  
Web site: <http://www.genoscope.cns.fr/>  
Contact: SeqRef@genoscope.cns.fr

The following BAC sequence is oriented from the 7' to the 3' end. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . . even if efforts are made to eliminate these contaminating sequences.

```

-----
Assembly program: Phrap, version 2.0
Quality coverage: 5.91x in Q20 bases; sum-of-contigs
-----
Summary statistics
-----
Upstream BAC (overlapping the 7' end) : C-3112B5
Downstream BAC (overlapping the 3' end) : R-346L24 (AC-A1163612)
-----

```



Contigs composition :  
 98062 bp contig from 1 to 98062  
 60765 bp contig from 98163 to 158927.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 98062: contig of 98062 bp in length  
 \* 98063 98162: gap of 100 bp  
 \* 98163 158927: contig of 60765 bp in length.  
 Location/Qualifiers  
 1. 158927  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /clone\_11b="RPC1-11"  
 77210..77462  
 /note="matching EMBL:Z66661  
 RHdb:RH53692  
 RHdb:RH53795  
 dbSTS:STS8872  
 Identified using the e-PCR software (G. Schuler)"  
 89489..89715  
 /note="matching EMBL:Z23897  
 RHdb:RH53481  
 RHdb:RH53636  
 dbSTS:STS9781  
 Identified using the e-PCR software (G. Schuler)"  
 89541..89698  
 /note="matching EMBL:Z23897  
 RHdb:RH95620  
 RHdb:RH49080  
 RHdb:RH13032  
 RHdb:RH1093  
 dbSTS:STS1260  
 Identified using the e-PCR software (G. Schuler)"  
 BASE COUNT 49681 a 32396 c 30501 g 46179 t 170 others  
 ORIGIN

Query Match 16.3% Score 20; DB 2; Length 158927;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 gatactatctcaattat 112  
 |||||  
 Db 148122 GATACCTATCTCAATTAT 148103

RESULT 8  
 CNS01R11/c 171345 bp DNA PRI 28-APR-2001  
 LOCUS Human chromosome 14 DNA sequence BAC R-346124 of library RPC1-11  
 DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.  
 ACCESSION AL163612 GI:13276126  
 VERSION  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 171345)  
 Hellig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,  
 Brothier,P., Catolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,  
 Levy,M., Eckenberg,R., Bruls,T., deBardins,V., Cruaud,C.,  
 Gysay,G., Saurin,W. and Weissendach,J.  
 Sequencing of the human chromosome 14

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 171345)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 On Mar 12, 2001 this sequence version replaced g1:11875958.  
 COMMENT  
 Center: Genoscope / Centre National de Sequencage  
 Center code: GS  
 Web site: http://www.genoscope.cns.fr/  
 Contact: seqref@genoscope.cns.fr  
 -----  
 The following BAC sequence is oriented from the T7 to the SP6 end.  
 Upstream BAC (overlapping the T7 end) : R-326E7  
 Downstream BAC (overlapping the SP6 end) : R-649E7 (AC-AL139099)  
 ----- Summary Statistics -----  
 Assembly program: Phrap; version 2.0  
 Quality coverage: 9.09x in Q20 bases; sum-of-contigs  
 -----  
 Overall quality chart :  
 Range : bases  
 0 :  
 1 - 9 :  
 10 - 19 : 2  
 20 - 29 : 9  
 30 - 39 : 186  
 40 - 49 : 2239  
 50 - 59 : 7900  
 60 - 69 : 8553  
 70 - 79 : 12847  
 80 - 89 : 31205  
 90 - 99 : 108404  
 -----  
 Percentage of bases with a quality value >= 40 : 99 %.

FEATURES  
 source  
 1..171345  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /clone\_11b="RPC1-11"  
 BASE COUNT 55535 a 35560 c 32962 g 47288 t  
 ORIGIN

Query Match 16.3% Score 20; DB 9; Length 171345;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 gatactatctcaattat 112  
 |||||  
 Db 25550 GATACCTATCTCAATTAT 25531

RESULT 9  
 AC024417 182340 bp DNA HTG 12-MAR-2000  
 LOCUS Homo sapiens chromosome 4 clone RP11-221C21 map 4, \*\*\* SEQUENCING  
 DEFINITION IN PROGRESS \*\*\*; 64 unordered pieces.  
 ACCESSION AC024417  
 VERSION AC024417.2 GI:7229920  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 182340)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 4, clone RP11-221C21  
 Unpublished



REFERENCE  
AUTHORS

2 (bases 1 to 182340)  
Birren,B., Linton,J., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,  
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,  
Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Kartas,A.,  
Klein,J., Landers,T., Laroque,K., Lehoczeky,J., Levine,R.,  
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,  
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,  
Meneus,L., Mihova,T., Miranda,C., Mienga,Y., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,  
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,  
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Sudramanlan,A., Talamas,J., Testaye,S., Theodore,J., Titrrell,A.,  
Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,  
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
Zody,M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 12, 2000 this sequence version replaced gi:7108213.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L6906

Center clone name: 221\_C\_21

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 64 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1056 1155: contig of 1055 bp in length  
1156 2524: contig of 1369 bp in length  
2525 2624: gap of 100 bp  
2625 3768: contig of 1144 bp in length  
3769 3868: gap of 100 bp  
3869 4906: contig of 1038 bp in length  
4907 5006: gap of 100 bp  
5007 6012: contig of 1006 bp in length  
6013 6112: gap of 100 bp  
6113 7527: contig of 1415 bp in length  
7528 7627: gap of 100 bp  
7628 8665: contig of 1238 bp in length  
8666 8965: gap of 100 bp  
8966 9989: contig of 1024 bp in length  
9990 10089: gap of 100 bp  
10090 11332: contig of 1243 bp in length  
11333 11432: gap of 100 bp  
11433 12597: contig of 1165 bp in length  
12598 12697: gap of 100 bp  
12698 13979: contig of 1282 bp in length  
13980 14079: gap of 100 bp  
14080 15544: contig of 1465 bp in length  
15545 15644: gap of 100 bp  
15645 17090: contig of 1446 bp in length  
17091 17190: gap of 100 bp  
17191 18508: contig of 1318 bp in length

18509 18608: gap of 100 bp  
18609 19829: contig of 1221 bp in length  
19830 19929: gap of 100 bp  
19930 21213: contig of 1284 bp in length  
21214 21313: gap of 100 bp  
21314 22808: contig of 1495 bp in length  
22809 22908: gap of 100 bp  
22909 24534: contig of 1626 bp in length  
24535 24634: gap of 100 bp  
24635 27162: contig of 2528 bp in length  
27163 27262: gap of 100 bp  
27263 28811: contig of 1549 bp in length  
28812 28911: gap of 100 bp  
28912 30046: contig of 1135 bp in length  
30047 30146: gap of 100 bp  
30147 31304: contig of 1158 bp in length  
31305 31404: gap of 100 bp  
31405 33474: contig of 2070 bp in length  
33475 33574: gap of 100 bp  
33575 35085: contig of 1511 bp in length  
35086 35185: gap of 100 bp  
35186 36846: contig of 1661 bp in length  
36847 36946: gap of 100 bp  
36947 38786: contig of 1840 bp in length  
38787 38886: gap of 100 bp  
38887 41015: contig of 2129 bp in length  
41016 41115: gap of 100 bp  
41116 43401: contig of 2286 bp in length  
43402 43501: gap of 100 bp  
43502 44790: contig of 1289 bp in length  
44791 44890: gap of 100 bp  
44891 46931: contig of 2041 bp in length  
46932 47031: gap of 100 bp  
47032 50023: contig of 2992 bp in length  
50024 50123: gap of 100 bp  
50124 52810: contig of 2687 bp in length  
52811 52910: gap of 100 bp  
52911 55013: contig of 2103 bp in length  
55014 55113: gap of 100 bp  
55114 57694: contig of 2581 bp in length  
57695 57794: gap of 100 bp  
57795 60148: contig of 2354 bp in length  
60149 60248: gap of 100 bp  
60249 62576: contig of 2328 bp in length  
62577 62676: gap of 100 bp  
62677 64748: contig of 2072 bp in length  
64749 64848: gap of 100 bp  
64849 67425: contig of 2577 bp in length  
67426 67525: gap of 100 bp  
67526 70617: contig of 3092 bp in length  
70618 70717: gap of 100 bp  
70718 72665: contig of 1948 bp in length  
72666 72765: gap of 100 bp  
72766 75102: contig of 2337 bp in length  
75103 75202: gap of 100 bp  
75203 78445: contig of 3243 bp in length  
78446 78545: gap of 100 bp  
78546 81397: contig of 2852 bp in length  
81398 81497: gap of 100 bp  
81498 84343: contig of 2846 bp in length  
84344 84443: gap of 100 bp  
84444 86480: contig of 2037 bp in length  
86481 86580: gap of 100 bp  
86581 88846: contig of 2266 bp in length  
88847 88946: gap of 100 bp  
88947 93512: contig of 4566 bp in length  
93513 93612: gap of 100 bp  
93613 96846: contig of 3234 bp in length  
96847 96946: gap of 100 bp  
96947 99557: contig of 2611 bp in length  
99558 99657: gap of 100 bp  
99658 102370: contig of 2713 bp in length  
102371 102470: gap of 100 bp



```

* 102471 106107: contig of 3637 bp in length
* 106108 106207: gap of 100 bp
* 106208 109111: contig of 2904 bp in length
* 109112 109211: gap of 100 bp
* 109212 113450: contig of 4239 bp in length
* 113451 113550: gap of 100 bp
* 113551 118394: contig of 4844 bp in length
* 118395 118494: gap of 100 bp
* 118495 123220: contig of 4726 bp in length
* 123221 123320: gap of 100 bp
* 123321 127732: contig of 4412 bp in length
* 127733 127832: gap of 100 bp
* 127833 132756: contig of 4924 bp in length
* 132757 132856: gap of 100 bp
* 132857 137254: contig of 4398 bp in length
* 137255 137354: gap of 100 bp
* 137355 143801: contig of 6447 bp in length
* 143802 143901: gap of 100 bp
* 143902 148857: contig of 4956 bp in length
* 148858 148957: gap of 100 bp
* 148958 156185: contig of 7228 bp in length
* 156186 156285: gap of 100 bp
* 156286 164408: contig of 8123 bp in length
* 164409 164508: gap of 100 bp
* 164509 171697: contig of 7189 bp in length
* 171698 171797: gap of 100 bp
* 171798 182340: contig of 10543 bp in length.

```

FEATURES  
source  
1. 182340  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"

misc\_feature  
1. 1055  
/clone="RP11-221C21"  
/clone\_lib="RPC1-11 Human Male BAC"

misc\_feature  
1156. 2524  
/note="assembly\_fragment"

misc\_feature  
2625. 3768  
/note="assembly\_fragment"

Query Match 16.3%: Score 20; DB 2; Length 182340;  
Best Local Similarity 100.0%; Pred.No. 1.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 ctatcttcaattattcca 116  
|||||

Db 178749 CTTATCTCAATTATTCCA 178768

RESULT 10  
AP001675/c 340000 bp DNA PRI 30-MAY-2000  
LOCUS Homo sapiens genomic DNA, chromosome 21q, section 19/105.  
DEFINITION AP001675 AL163220 BA000005  
ACCESSION  
VERSION AP001675.1 GI:7768691  
KEYWORDS  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (sites)  
REFERENCE  
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,  
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,  
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,  
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,  
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.  
The DNA sequence of human chromosome 21. The chromosome 21 mapping  
and sequencing consortium  
Nature 405. (6784), 311-319 (2000)

JOURNAL  
MEDLINE  
20289799  
2 (bases 1 to 340000)  
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,

# TITLE JOURNAL

Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,  
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,  
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,  
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,  
Zimmerman,M., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,  
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,  
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,  
Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,  
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,  
Hennig,S., Risseilmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,  
Gardner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and  
Vaspo,M.L.

## COMMENT

Direct Submission  
Submitted (10-Apr-2000) to the DDBJ/EMBL/GenBank databases. The  
Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic  
Sciences Center, Human Genome Research Group \* Institute of  
Molecular Biotechnology, Genome Analysis \* Keio University School  
of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome  
Analysis \* Max-Planck Institute for Molecular Genetics (addresses  
see below)  
On May 30, 2000 this sequence version replaced gi:7717272.  
The chromosome 21 mapping and sequencing consortium consisting of  
\* RIKEN Genomic Sciences Center, Human Genome Research Group, \*  
Sagamihara 228-8555, Japan,  
\* e.mail: hattori@gscl.riken.go.jp/  
\* URL: http://hgp.gscl.riken.go.jp/  
and  
\* Institute of Molecular Biotechnology, Genome Analysis, \*  
Beutenbergstrasse 11, D-07745 Jena, Germany,  
\* e.mail: gscj-submit@genome.imb-jena.de  
\* URL: http://genome.imb-jena.de/  
and  
\* Keio University School of Medicine, Molecular Biology, \* Tokyo  
160-8582, Japan,  
\* e.mail: nshimiz@dm.med.keio.ac.jp/  
\* URL: http://www.dmb.med.keio.ac.jp/  
and  
\* GBF, Dept. of Genome Analysis,  
\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e.mail:  
info.genome@gbf.de  
\* URL: http://genome.gbf.de/  
and  
\* Max-Planck Institute for Molecular Genetics,  
\* Innesstrasse 73, D-14195 Berlin, Germany,  
\* e.mail: info-chr21@molgen.mpg.de  
\* URL: http://chr21.rz-berlin.mpg.de/  
AL163220: Submitted (10-Apr-2000).

## FEATURES

source

source

source

source

repeat\_region

repeat\_region

1. 340000  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1. 9371  
/note="Accession No. AL132993"  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q21.1"  
/clone="c103E2331, 5' partial"  
/clone\_lib="ICRFc103 Cosmid Library"  
<1. 113120  
/note="Accession No. AP000431"  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q21.1"  
/clone="R28W9, 5' partial"  
/clone\_lib="RPC1-11 BAC library"  
complement(1510..1686)  
/note="LIMD1"  
/rpt\_family="LINE/L1"  
/rpt\_type="DISPERSED"  
complement(1719..1947)



```

/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region complement(1948. .2340)
/note="L1M2"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region 4085. .4116
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region 4117. .4158
/note="(TTTG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region 4139. .4309
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region 4336. .5118
/note="L1P5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region complement(5588. .6010)
/note="MLT1H"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 6279. .6642
/note="MSTA"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 6648. .7011
/note="MSTA-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region complement(7020. .11228)
/note="L1Pa2"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region 11194. .11962
/note="MSTA-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 11963. .12333
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 13901. .14287
/note="THE1A"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 14288. .14493
/note="MSTA-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 14494. .15181
/note="LTR8"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
repeat_region 15182. .15519
/note="MSTA-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 15523. .15620
/note="MSTA"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region complement(15621. .16036)
/note="MSTA"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 16091. .16223
/note="L1MA1"

```

```

/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region 16256. .16529
/note="MSTA"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 17192. .17214
/note="(TTTGA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region complement(17215. .17497)
/note="Alusg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region complement(18031. .18355)
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 18356. .18783
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region 18946. .19195
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region 19247. .19419
/note="LTR35A"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
repeat_region 19462. .19540
/note="Alusg/x"
/rpt_family="SINE/Alu"

```

Query Match 16.3%; Score 20; DB 9; Length 340000;  
 Best Local Similarity 100.0%; Pred.No. 1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 ctatcttcaattatcca 116  
 |||||||  
 Db 329060 CTTATCTCAATTATTCa 329041

```

RESULT 11
LOCUS      U41746      27710 bp      DNA      INV      10-JUL-2001
DEFINITION Caenorhabditis elegans cosmid T18H9, complete sequence.
ACCESSION  U41746
VERSION    U41746.1 GI:1118132
KEYWORDS   HTG.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans.
REFERENCE  1 (bases 1 to 27710)
AUTHORS    The C. elegans Sequencing Consortium.
TITLE      Investigating biology. The C. elegans Sequencing Consortium
JOURNAL    Science 282 (5396), 2012-2018 (1998)
MEDLINE    99069613
REFERENCE  2 (bases 1 to 27710)
AUTHORS    Gattung,S. and Du,Z.
TITLE      The sequence of C. elegans cosmid T18H9
JOURNAL    Unpublished
AUTHORS    Waterston,R.
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  4 (bases 1 to 27710)
AUTHORS    Waterston,R.
TITLE      Direct Submission
JOURNAL    Submitted (30-NOV-1995) Robert Waterston
REFERENCE  5 (bases 1 to 27710)

```







```
VSLSBQLCSILGSSMSPEGSDAWVLCNLSMPAYSLIGGPVPIVNCPSFQVOTLISH
VNIIONFCHMSNSNPETIVGILGDTDFLSOVKAYVYVLCNLSMPAYSLIGGPVPIVNCPSFQVOTLISH
PSTSLFYKMGICIDPOLNLCRDLMRDCMDTLPEKTSILANKSIAMPNVSKSKNM
PIGEAMQITADBDPEGGRVDFLSEVAVGSOQCEIENDDDNGINGRQVQVITPRE
ESVISTKTDGSSPHSPQLRNVDOELMDPVPNNPPOSSQSYSGSDPSPSSNTNPN
SKKDTAKVTKIKAFKTLVSRSPSSGLSLTYLKEKRMKMLQKMGKQKQKPEEG
AAPSQVSSIGRMICATGKHNELTVIIDGNAYNGIRYFQTSQWQMHKISPIAFETP
INCTA"
gene
25268. 27631
/ gene="T18H9.1"
/ note="gid-6"
join(25268. .25334,25502. .25749,25795. .26001,26180. .26566,
26617. .26915,26956. .27037,27086. .27411,27487. .27631)
/ gene="T18H9.1"
/ note="coded for by the following C. elegans CDNAs:
yk68a8.5"
/ codon_start=1
/ product="Hypothetical protein T18H9.1"
/ protein_id="AAA83334.1"
/ db_xref="GI:118137"
/ translation="MPSIRYRLVALVVFISVYGOEAAVQTPPVQPNLIRGSSG
EPVRAIRPNQIOLPDIDPLRDERKRNHAGYRHPTFQVPOQOQFRPAGVY
IAQPNRVYSQTFRSAPVRFVIFQOQFQOGOSQITQOQIRQOQOQSQTRFDL
RQQVAPRVAVYIAGRLTGPIAVYQGSERPAVAVIERPVAPRTPIERPV
ARPAPIERPEPAPAPYIERPVAPAPYIERPVAPAPYIERPVAPAPYIERPV
RPVAPSTTORIETPTQPTTRATRTKRTTTTAAPTPRLTATATPLATTSR
```

```
Query Match 15.4%; Score 19; DB 3; Length 27710;
Best Local Similarity 100.0%; Pred.No. 5.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 98 ttatctcaattattcca 116
|||||
Db 23408 TTATCTTCATTTATTCCA 23390
```

```
RESULT 12
AL357912
LOCUS Human DNA sequence from clone RP11-328N1 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL357912 AC022524
VERSION AL357912.10 GI:14018273
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 130057)
AUTHORS Hall, R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On May 15, 2001 this sequence version replaced gi:8569932
gi:13751415.
```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TrEMBL; Wp; WormPep; Information on the WormPep database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP11-328N1 is from the library RP11-328N1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-328N1. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-328N1 is at 1 in this sequence. The true left end of clone RP5-1102M4 is at 129958 in this sequence.

## FEATURES

```
source
1..130057
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-328N1"
/clone_1lb="RPCT-11.2"
3..166
/ note="MSTC repeat: matches 1..168 of consensus"
171..1394
/ note="LM1C repeat: matches 321..1600 of consensus"
1431..1870
/ note="LM4 repeat: matches 3258..3689 of consensus"
1871..2174
/ note="AlusJ repeat: matches 1..306 of consensus"
2175..3813
/ note="LM4 repeat: matches 3689..5371 of consensus"
3814..8354
/ note="LM4 repeat: matches 1499..6145 of consensus"
8355..8676
/ note="LM1 repeat: matches 5814..6123 of consensus"
8685..8990
/ note="AlusX repeat: matches 1..291 of consensus"
9080..9234
/ note="HAI1 repeat: matches 1523..1692 of consensus"
10201..10493
/ note="AlusX repeat: matches 1..293 of consensus"
10626..10923
/ note="AluJb repeat: matches 1..297 of consensus"
12090..12320
/ note="MER20 repeat: matches 1..218 of consensus"
13930..14603
/ note="LM1A3 repeat: matches 5620..6304 of consensus"
14617..20083
/ note="LM1A4 repeat: matches 649..6139 of consensus"
20157..20374
/ note="L1R9B repeat: matches 427..644 of consensus"
20376..20630
/ note="LM1A3 repeat: matches 5376..5630 of consensus"
20631..21180
/ note="MLT2B repeat: matches 1..448 of consensus"
21183..27754
/ note="LM1A3 repeat: matches -1387..5378 of consensus"
27850..27996
/ note="LM1 repeat: matches 238..386 of consensus"
28485..29096
/ note="L2 repeat: matches 2064..2709 of consensus"
29211..29319
/ note="L2 repeat: matches 1687..1795 of consensus"
29487..29592
/ note="L2 repeat: matches 1335..1441 of consensus"
30157..30218
/ note="31 copies 2 mer tt 67% conserved"
31310..31624
/ note="Alus repeat: matches 1..311 of consensus"
32682..33408
/ note="L1R8 repeat: matches 1..691 of consensus"
33523..33812
/ note="LM4 repeat: matches 3952..4265 of consensus"
```



```

repeat_region 33829..34304
/note="L1MA2 repeat: matches 5815..6289 of consensus"
repeat_region 34305..34552
/note="AlusX repeat: matches 38..285 of consensus"
repeat_region 34556..34605
/note="25 copies 2 mer aa 72% conserved"
repeat_region 36334..36494
/note="TIGER1 repeat: matches 2278..2418 of consensus"
repeat_region 36495..36761
/note="AlusX repeat: matches 1..267 of consensus"
repeat_region 36762..38999
/note="TIGER1 repeat: matches 49..2278 of consensus"
repeat_region 39000..40970
/note="L1PA3 repeat: matches 1596..3569 of consensus"
repeat_region 40964..43533
/note="L1PA3 repeat: matches 3578..6146 of consensus"
repeat_region 43561..43620
/note="TIGER1 repeat: matches 1..60 of consensus"
repeat_region 43899..44415
/note="MTD repeat: matches 1..505 of consensus"
repeat_region 44915..50056
/note="L1PA1 repeat: matches 951..6163 of consensus"
repeat_region 50057..50356
/note="AlusX repeat: matches 9..308 of consensus"
repeat_region 50357..50428
/note="L1PA1 repeat: matches 880..951 of consensus"
repeat_region 50410..51676
/note="L1PA1 repeat: matches -1422..750 of consensus"
repeat_region 52792..53029
/note="L2 repeat: matches 1740..1982 of consensus"
repeat_region 53427..53456
/note="15 copies 2 mer ta 86% conserved"
repeat_region 53883..54161
/note="MER44A repeat: matches 57..330 of consensus"
repeat_region 55356..55645
/note="AlusX repeat: matches 1..293 of consensus"
repeat_region 57603..57720
/note="59 copies 2 mer ag 71% conserved"
repeat_region 59372..59459
/note="MER91B repeat: matches 90..180 of consensus"
repeat_region 60347..61105
/note="L1PA3 repeat: matches 5997..6156 of consensus"
repeat_region 61106..61408
/note="AlusX repeat: matches 1..303 of consensus"
repeat_region 61409..63432
/note="L1PA3 repeat: matches 3326..5397 of consensus"
repeat_region 63446..63718
/note="AlusX repeat: matches 3..292 of consensus"
repeat_region 65907..66088
/note="MER5A repeat: matches 1..183 of consensus"
repeat_region 67083..67456
/note="L1 repeat: matches 4640..5052 of consensus"
repeat_region 67669..67874
/note="L1PA4 repeat: matches 5319..5531 of consensus"
repeat_region 68960..69182
/note="MTD repeat: matches 234..466 of consensus"
repeat_region 69306..69415
/note="MTD repeat: matches 124..226 of consensus"
repeat_region 69416..69723
/note="AlusX repeat: matches 1..307 of consensus"
repeat_region 69724..69811
/note="MTD repeat: matches 16..124 of consensus"
repeat_region 70639..70787
/note="L1MA6 repeat: matches 6027..6183 of consensus"
repeat_region 73464..73779
/note="L1MA4 repeat: matches 2938..3264 of consensus"
repeat_region 73790..75563
/note="L1MA2 repeat: matches 4311..6108 of consensus"
repeat_region 75564..75876
/note="AlusX repeat: matches 1..312 of consensus"
repeat_region 75877..76621
/note="L1MA2 repeat: matches 3554..4311 of consensus"
repeat_region 78308..78593

```

```

repeat_region /note="AlusB repeat: matches 1..287 of consensus"
80898..86972
/note="L1PA4 repeat: matches 11..6144 of consensus"
repeat_region 86996..88410
/note="L1 repeat: matches 3108..4548 of consensus"
repeat_region 88411..88491
/note="Alus repeat: matches 222..302 of consensus"
repeat_region 88499..88549
/note="THE1B repeat: matches 136..195 of consensus"
repeat_region 88550..88762
/note="AlusX repeat: matches 1..216 of consensus"
repeat_region 88763..89267
/note="L1 repeat: matches 2527..3102 of consensus"
repeat_region 89269..89605
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 89606..90914
/note="THE1B-INTERNAL repeat: matches 1..1354 of consensus"
repeat_region 90916..91268

```

Query Match 15.4%; Score 19; DB 9; Length 130057;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 ctatctcaattatctcc 115  
 Db 77682 CTTATCTCAATTATTC 77700

RESULT 13  
 AC027498/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome 18 clone RP11-656L11 map 18, WORKING DRAFT  
 SEQUENCE, 25 unordered pieces.  
 AC027498  
 AC027498.2 GI:7622437  
 VERSION  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE  
 ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 180911)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Bouckgalter,B., Brown,A., Buxfelt,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Glende,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,U., Lacroque,K., Lamazares,R., Landers,T., Lehotzky,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,  
 Meldrum,J., Menus,L., Mihova,C., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Olivat,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testfay,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
 JOURNAL  
 COMMENT  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)



```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8451
Center clone name: 656.L11
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165132 bases at least Q40
Consensus quality: 173168 bases at least Q30
Consensus quality: 176502 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 178511; sum-of-ctrls
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
2315: contig of 2315 bp in length
2316 2415: gap of 100 bp
2416 6148: contig of 3733 bp in length
6149 6248: gap of 100 bp
6249 9065: contig of 2817 bp in length
9066 9165: gap of 100 bp
9166 13447: contig of 4282 bp in length
13448 13547: gap of 100 bp
13548 17515: contig of 3968 bp in length
17516 17615: gap of 100 bp
17616 21811: contig of 4196 bp in length
21812 21911: gap of 100 bp
21912 26026: contig of 4115 bp in length
26027 26126: gap of 100 bp
26127 29586: contig of 3460 bp in length
29587 29686: gap of 100 bp
29687 34675: contig of 4989 bp in length
34676 34775: gap of 100 bp
34776 39152: contig of 4377 bp in length
39153 39252: gap of 100 bp
39253 44142: contig of 4890 bp in length
44143 44242: gap of 100 bp
44243 48561: contig of 4319 bp in length
48562 48661: gap of 100 bp
48662 53987: contig of 5326 bp in length
53988 54087: gap of 100 bp
54088 60371: contig of 6284 bp in length
60372 60471: gap of 100 bp
60472 67910: contig of 7439 bp in length
67911 68010: gap of 100 bp
68011 74138: contig of 6128 bp in length
74139 74238: gap of 100 bp
74239 80142: contig of 5904 bp in length
80143 80242: gap of 100 bp
80243 88060: contig of 7818 bp in length
88061 88160: gap of 100 bp
88161 95098: contig of 6938 bp in length
95099 95198: gap of 100 bp
95199 104637: contig of 9439 bp in length
104638 104737: gap of 100 bp
104738 114983: contig of 10246 bp in length
114984 115083: gap of 100 bp
115084 124831: contig of 9748 bp in length
124832 124931: gap of 100 bp

```

```

FEATURES
Source
124932 136857: contig of 11926 bp in length
136858 136957: gap of 100 bp
136958 150860: contig of 13903 bp in length
150861 150960: gap of 100 bp
150961 180911: contig of 29951 bp in length.
Location/Qualifiers
1. 180911
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-656L11"
/clone_11p="RPC1-11 Human Male BAC"
1. 2315
/note="assembly-fragment"
2416. 6148
/note="assembly-fragment"
6249. 9065
/note="assembly-fragment"
9166. 13447
/note="assembly-fragment"
13548. 17515
/note="assembly-fragment"
17616. 21811
/note="assembly-fragment"
21912. 26026
/note="assembly-fragment"
26127. 29586
/note="assembly-fragment"
29687. 34675
/note="assembly-fragment"
34776. 39152
/note="assembly-fragment"
39253. 44142
/note="assembly-fragment"
44243. 48561
/note="assembly-fragment"
48662. 53987
/note="assembly-fragment"
54088. 60371
/note="assembly-fragment"
60472. 67910
/note="assembly-fragment"
68011. 74138
/note="assembly-fragment"
74239. 80142
/note="assembly-fragment"
80243. 88060
/note="assembly-fragment"
88161. 95098
/note="assembly-fragment"
95199. 104637
vector_side:left"
104738. 114983
/note="assembly-fragment"
115084. 124831
/note="assembly-fragment"
124932. 136857
/note="assembly-fragment"
clone_end:Spe
vector_side:right"
136958. 150860
/note="assembly-fragment"
150961. 180911
/note="assembly-fragment"
BASE COUNT 58145 a 33104 c 32922 g 54339 t 2401 others
ORIGIN
Query Match 15.4%; score 19; DB 2; length 180911;
Best Local Similarity 100.0%; Pred. No. 4.1;

```







[illegible]

```

CDS      /gene="NMB0397"  

         8685.. 8972  

         /gene="NMB0397"  

         /note="hypothetical protein; identified by Glimmer2;  

         putative"  

         /transl_table=1  

         /codon_start=1  

         /product="hypothetical protein"  

         /protein_id="AAF40837.1"  

         /db_xref="GI:7225619"  

         /translation="MSGKEPLNPMPVINSGRKCRDKSSDIFRKGDALEKTAFQAQAA  

         ILFRQGIGIRACKRTGCGSFMFSESPFLPAKKVKGGYINPVOSSINI"  

         9064.. 9345  

         /gene="NMB0398"  

         9064.. 9345  

         /gene="NMB0398"  

         /note="Similar to GB:U00096 PID:1789021 PID:1800053  

         PID:1800057 percent identity: 59.26; Identified by  

         sequence similarity; putative"  

         /codon_start=1  

         /transl_table=1  

         /product="transcriptional regulator, ArsR family"  

         /protein_id="AAF40838.1"  

         /db_xref="GI:7225620"  

Query Match 14.6%; Score 18; DB 1; Length 11018;  

Best Local Similarity 100.0%; Pred. No. 21;  

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.  

Oy 97 ctatcttcacatttatc 114  

|||||  

Db 10210 CTATCTTCACATTATTC 10227  

RESULT 16  

LOCUS CELC34H3 33803 bp DNA INV 28-MAR-2000  

DEFINITION Caenorhabditis elegans cosmid C34H3, complete sequence.  

ACCESSION AC006614  

VERSION AC006614.1 GI:4263324  

KEYWORDS HTG.  

SOURCE Caenorhabditis elegans.  

ORGANISM Caenorhabditis elegans  

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;  

Rhabditoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.  

REFERENCE 1 (bases 1 to 33803)  

AUTHORS The C. elegans Genome Sequencing Consortium, Washington University  

Genome Sequencing Center, St. Louis U.S.A. and The Sanger Centre,  

Hinixon, U.K.,C.  

TITLE Genome sequence of the nematode C. elegans: a platform for  

investigating biology. The C. elegans Sequencing Consortium  

99069613 Science 282 (5396), 2012-2018 (1998)  

ERRATUM:[published errata appear in Science 1999 Jan  

1,283(5388):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep  

3,285(5433):1493]]  

JOURNAL MEDLINE 2 (bases 1 to 33803)  

REMARK Lamar,B. and Wamsley,P.  

REFERENCE The sequence of C. elegans cosmid C34H3  

AUTHORS Unpublished  

TITLE 3 (bases 1 to 33803)  

JOURNAL Waterston,R.H.  

REFERENCE Direct Submission  

AUTHORS Submitted (23-FEB-1999) Genome Sequencing Center, Washington  

TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  

JOURNAL MO 63108, USA  

REFERENCE 4 (bases 1 to 33803)  

AUTHORS Waterston,R.H.  

TITLE Direct Submission  

JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington  

TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  

JOURNAL MO 63108, USA  

REFERENCE 5 (bases 1 to 33803)  


```



AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 6 (bases 1 to 33803)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

## COMMENT

Submitted by:

Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 e-mail: twenematode.wustl.edu and jesusanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

## NEIGHBORING COSMID INFORMATION:

The 5' clone is Y72A10A, 200 bp overlap; 3' clone is T25B2, 200 bp overlap. Actual start of this clone is at base position 1 of CELC34H3; actual end is at 33607 of CELC34H3

## NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder(P. Green and L. Hallier, ms in preparation).

## FEATURES

## source

1..33803

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/chromosome="X"

/clone="C34H3"

/complement(8252..16090)

/gene="C34H3.1"

/complement(join(8252..8406,8464..8572,9494..9597,9653..9756,9809..9894,9946..10146,13178..13307,13542..13628,14243..14403,14887..14998,15046..15146,15195..15266,15855..15995,16043..16090))

/gene="C34H3.1"

/note="contains similarity to Pfam family PF00200 (Disintegrin), score=15.2, E=0.00052, N=1; coded for by C. elegans cDNA yk228a1.5; coded for by C. elegans cDNA yk344b12.5; coded for by C. elegans cDNA yk228a1.3; coded for by C. elegans cDNA yk344b12.3"

/codon\_start=1

/protein\_id="AAF39762.1"

/db\_xref="GI:7206602"

/translation="MRKSLILVFCIFPNVSTLEQKHEYGAVGAKADYSIAS

TDSGTVFIRSLVFDNKTATYEDMTIVKLNIMKAVDEANQYLQDLVGLIVGIL

QTRGDLDSQSFHEYNRLKLPHEFTLLSYKAGGLAVNGCSSHSVLSGFY

PNEPRAMGSIFFHEVAHLGVPHRAVNESIYVPCIPKDSLKEGCKIKIGFDHC

TVOQFVNTLYKKNKILKHPILFQSEVPCGNGVLENDGDCGPGCSLQNPCHRC

FEMHPFLVLIVASFVAFVLTWFTVRYGTGTMNCEMYSKSHDRGNASPTNCOI

QILASPYONRKMSHSISGNTILVANSRATORRKPVPVPPPKRTTIOVAVGS

GOYETTRVFGSYRESFYDDSDDEDEDEVSATPLPGVPTCPAYPNPNNRYTO

CLLPQNTSHCAPSSHQTFHFDAGAVLNIHMSILSFQGYVDMLFSAHCFEFLYE

## gene

LVLPFFNFELKTTFTTLLNLIISPCR"

## gene

complement(25250..26090)

## gene

/gene="C34H3.2"

## gene

complement(join(25250..25703,25780..26090))

## gene

/gene="C34H3.2"

## gene

/note="contains similarity to Pfam family PF00096 (zf-C2H2), score=77.4, E=3e-19, N=3"

## gene

/codon\_start=1

## gene

/evidence="not\_experimental"

## gene

/protein\_id="AAF39763.1"

## gene

/db\_xref="GI:7206603"

## gene

/translation="MLPMORVPTSIPOSNQVFRMLAQHQLQNLQORRMALL

## gene

AMNPEIPMTDLAKKAFEDTHADSISSQKIKESVSPKSPITTTAAVPRPYDQ

## gene

PWMIPGRGRTTGRRARPKERTICKDHFHTSTNLIHERHTDERPISCDVCGRA

## gene

FRQDRLRDHKYIHKDRPKCEICGCGOSRTLLVHRATHDPNRHSIGAVVPKIS

## gene

ETPILPELDPRVLLIQLNLDNSFNSTSPQISPDPR"

## BASE COUNT

11637 a 5470 c 6333 g 10363 t

## ORIGIN

## Query Match

14.6%; Score 18; DB 3; Length 33803;

## Best Local Similarity

100.0%; Pred. No. 18;

## Matches

18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## Qy

99 tatcttaattattcca 116

## Db

4201 TATCTTCAATTATTCOA 4184

## RESULT 17

## LOCUS

AC022128 34128 bp DNA PRI 30-MAR-2001

## DEFINITION

Homo sapiens chromosome 5 clone CTD-2170613, complete sequence.

## ACCESSION

AC022128

## VERSION

AC022128.5 GI:13489154

## KEYWORDS

HTG.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

## TITLE

DOE Joint Genome Institute and Stanford Human Genome Center.

## JOURNAL

Direct Submission

## REFERENCE

2 (bases 1 to 34128)

## AUTHORS

DOE Joint Genome Institute.

## TITLE

Direct Submission

## JOURNAL

Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

## REFERENCE

3 (bases 1 to 34128)

## AUTHORS

DOE Joint Genome Institute and Stanford Human Genome Center.

## TITLE

Direct Submission

## JOURNAL

Submitted (30-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

## COMMENT

On Mar 30, 2001 this sequence version replaced gi:7711703.

## REFERENCE

www.fgi.doe.gov

## AUTHORS

Draft Sequence Produced by DOE Joint Genome Institute

## TITLE

Finishing completed at Stanford Human Genome Center

## JOURNAL

www.shgc.stanford.edu

## REFERENCE

Quality: Phrap Quality >=40 99.98% of Sequence;

## AUTHORS

Estimated Total Number of Errors is 0.1.

## TITLE

Note: Consensus clipped at overlap with AC008790.

## JOURNAL

Location/Qualifiers

## FEATURES

1..34128

## source

/organism="Homo sapiens"

## BASE COUNT

10725 a 7079 c 6751 g 9573 t

## ORIGIN

/db\_xref="taxon:9606"

## Query Match

14.6%; Score 18; DB 9; Length 34128;



Best Local Similarity 100.0%; Pred. No. 18;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 caacctgatyacctaat 45  
|||||

Db 29094 CAACCTGATGACTTAAT 29111

RESULT 18  
AC078862 57000 bp DNA PRI 23-MAR-2001  
LOCUS Homo sapiens 12q BAC RP11-139B1 (Roswell Park Cancer Institute  
DEFINITION Human BAC library) complete sequence.  
AC078862  
AC078862.14 GI:13435185  
KEYWORDS  
SOURCE HTG.  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 57000)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alshrook,S.L., Amaralunge,H.C., Are,D.R., Banks,T., Barbria,J.,  
Benton,D., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chiu,D., Chowdhury,I., Christopoulos,C., Cleveland,C.D.,  
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davys,C.,  
Devay-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,  
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,  
Garcia,A., Garner,T., Garza,N., Gill,R., Girelli,J.H., Guevara,W.,  
Gunnarathne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,  
Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O.,  
Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsi,F.,  
Howard,S., Huber,J., Huylk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,  
Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,  
Meador,M., Mel,G., Merscher,S., Metzger,M., Miller,A., Miner,G.,  
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,  
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,  
Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shm,C.,  
Shooshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A.,  
Stanley,R., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tameris,A.,  
Tameris,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,  
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D.,  
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,  
Washington,C., Watlington,S., Williams,G., Williamson,A.,  
Wlecyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,  
Zorilla,S., Kuchelapatti,R. and Gibbs,R.

## JOURNAL

## COMMENT

Submitted (23-MAR-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Mar 23, 2001 this sequence version replaced gi:13173525.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig length: 156673  
Phrap values in estimate: 156142  
Average error rate (BCM-Phrap estimate): 4.07525e-05  
Fraction of Phrap values less than 40 : 0.00374019  
Number of consensus changing edits: 306  
Number of N's in consensus : 0

----- Consensus changing edits -----  
Position Original+Context Edited+Context  
36 aaaggtcctt(n)agtttcctt aaaggtcctt(g)agtttcctt  
6213 tctctgttt(n)tcacagttt tctctgttt(c)tcacagttt  
14543 tgcctcaccc(n)annnnnnnnn tgcctcaccc(t)agatgtaac  
14545 ctcacatcna(n)ntgnnnnnnnn ctcacatcna(g)atgtaacatc  
14546 tccatccan(n)tgannnnnnn tccatccan(a)tgatgtaacatc  
14549 atccanantg(n)nnnnnnccc atccanantg(t)gaatcacaac  
14550 tccannntgn(n)nnnnnnccc tccannntgn(g)aatcacaac  
14551 cccannntgn(n)annnncccn cccannntgn(a)aatcacaac  
14553 nannntgnna(n)nnnncccn nannntgnna(t)aatcacaac  
14554 annntgnnan(n)nncccnnt annntgnnan(c)atcaccctgt  
14555 ntgnnnann(n)nncccnntn ntgnnnann(a)atcaccctgt  
14556 nannnnann(n)cccnntn nannnnann(n)atcaccctgt  
14561 nannnnccc(n)ntnnnnat nannnnccc(t)atcaccctgt  
14563 nnnnnccnt(n)tnnnnatn nnnnnccnt(g)atcaccctgt  
14565 nnncccnnt(n)nnnnatnn nnncccnnt(c)atcaccctgt  
14566 ncccnntln(n)nnatnnnnn ncccnntln(a)atcaccctgt  
14567 ccccnntnn(n)natnnnncc ccccnntnn(g)atcaccctgt  
14568 cccntnnnn(n)natnnnncc cccntnnnn(a)atcaccctgt







TEGINTIAGVGVGASOLMNVNGFOOTSACQSESIPISSSLAHYSOSIPLNAAVLOLOS  
VPPSALHYPOSTIPGSSLOYPOSTIPGSSSYOYPOSTIPGSSASSYGYIPYOVYHVOH  
GERRRPLVYDSSSNYSAGETTSIPDIOGHVQOOGMAGCYPRGSTRPSTQALAE  
OKVSDMKIREVEPENRKTPTGNDHONPOIDIVEVKNHNVREMAVATPPSODALH  
LPPSRDPRONTAKPATYRDVAITGVPLSGIDQLSTSSSTYAPVHSDSNLIDLN  
YPRPEQSSORVCYCSERIPEOLELNLNLSKSDNSLSQFVTSSESPANTAOODSGKEV  
GKSHDEKTYVNDANHHTHKQVETFEKVGVSDTELESEPLHKIVNDPDKNNRVV  
ADREIGVSNLSHVNAAMSHVTPPEQASLOGDILIDIDRPPRDLSEIFESQALSEDT  
TVRYPHSDAAVSMNVONHNRKMSYFOQALAEQFIORDVYVLODASRIPSDKDGSE  
SSRLPYVPSLRDGIINTNLNPOLITLQDYGNGSEKGGGTSIPALNEDMKATE  
SEEFGAVENTLRTPDSEPKDEKTETRAALPLGSEFDYSGLOIKNEDLELELGS  
GTFYVYHGKMGSDVAIKRIKSCFGRSSQDBRLGSEFGAEILISKJHNNVAV  
YGVYKDGCGGTATVTEYVYDGLRHVLYKDRDLDRKRLIITAMDAAFMKNGVSN  
TVHEDKCDNLVNLKDPSPRICKGVGFGSLKIKRNTLYSGVGRGTLPMMAPELNS  
SSKVEKGVDPESFGYIVMELTGEPEYANMHGALITGIGVYNNLRPIIPFCDDERT  
LMECHAPNMAPRSTETIAGRLRVHSSAATSTQSKPSAHRAK"

gene  
complement(5680..11012)  
/gene="TRK14.2"  
complement(join(5680..5757,5848..5928,6021..6128,  
6221..6295,6397..6516,6700..6876,6987..7103,7197..7316,  
7406..7443,7567..7787,7887..7999,8095..8262,8367..8507,  
8611..8853,8933..9010,9243..9520,9737..9963,10170..10354,  
10584..11012))  
/note="Is a member of P10004 ATPases associated with  
various cellular activities (AAA) family. ESTs gb|T43031,  
gb|R64750, gb|AA394742 and gb|A1100347 come from this  
gene."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAD30220.1"  
/db\_xref="GI:4835753"

CDS  
/translation="MEIAISYKPNPLISSSTOLLKRSKSGIVRPKAYEGATRRKO  
LRFVYASSSSSSSNDGFSWVRLAQSTRIGAEKIGESVYKTEIFDSEBASK  
RVMEIYARVADSVHKGHMLTRFKNETVPSFIDMKNMEHMKDIRMWDGKRVALLTA  
FALLSCQRYVAIQAPVERERELTESMEALIPESPNGINKEKERNMRAATPAG  
LKLRFTEADGTLVHDSYVGENAMDDLETGSLKILGNARIOTPAKKLSOD  
LGVSEIGDSVGNMRELTAKWELEKELSEOLNSSAKVYVEFMDEKESLREY  
IGRSETEGRALMSKRMWRYPKLPYTFYLOKIDSSAAVYVEFMDEKELRVITME  
GPLEYIVDIPLDYLPETICNAGVEVDILOKROIHEKVFALPGLITLIMFIRS  
AMLLITSRKFLTKTKNOLDMAYAEMLIPVGDVSTKSMKEVYVGGVWMDLDEL  
MIYMGNMQYERDVAVRGVLLSGPRTGKTLFARTLAKESGLPVGAGAEATPDE  
KSGAAKINEMESLARMAPAFVEDEIDAIAGRRADPRRRATFEALIDLOGEKX  
TGIDRESLRQAVIFICATNRDEDELEFVSGRRDRLYIGLPDOKORVOIFVHSG  
KNLAEDIDEGCANIRNLVNEAIIAMSVAKGSIYODIIVDLKOLLEGVNLTEEE  
QCKCEOSVSEYKRLILAVHAGHIVLAHLFPRPDMAFSOLLPGKRTANSVYPRD  
WVDGTTTETGMOMVAVHGGRCVAERVAGVDVTGGKDLERIKTITAREVITSPDS  
ARGLTQLVKRIQMVDPDMPDGLIKRYMDHHPVMEVSELSFTFELRYIEE  
TEELAMNALRANHILDLITRELLEKSRITGLVEEEMKDLSPLMEDFVQPIIND  
DEELPHKDRVSOPVDLRAAPLHRS"

gene  
11693..13641  
/gene="TRK14.3"  
join(11693..11887,12188..12265,12354..12614,12703..13020,  
13129..13287,13447..13641)  
/note="Is a member of the P100162 Phosphoglycerate kinase  
family. ESTs gb|N38721, gb|T22178, gb|B90345, gb|B90715,  
gb|T21440, gb|T46295, gb|H37082, gb|T46076, gb|N37132,  
gb|AA597649, gb|A1100648 and gb|Z48462 come from this  
gene."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAD30221.1"  
/db\_xref="GI:4835754"

CDS  
/translation="MAITKRSVGLKLEADLKGSVFWVDLNVPLDONSITDDTRIRA  
AVPTIKYLMGNSRVVLSHIGRPKGVTPYSIKPLVPRISLGLVEVNAVANDSICE  
VOKLVAGPEGVILLENVFAEENKDEPAKPLAALADVYVNDVAGFABRAHAST  
EGVAKFLKPSVAGELMOKELDYLVGAVANKRPFAIVGSKSVTKVGTESIILNPD  
IILGGMFTFEKKAQSLVSGSSVLEDEKIDLAKSLMEKKAQVSLILPTDIYIADK  
FAPDANSKIYPAITAPDGMGLDIGDPSIKTFSEALDITKTIITWNGMGVFEEDKRP  
GTEAVAKQALAEISGKGVYTIIGGSDSVAAEKVLADKMSHISTGGASLELLEGRPL  
PGVALDEA"

gene  
15312..17654

CDS  
/gene="TRK14.4"  
15312..17654  
/gene="TRK14.4"  
/note="Contains similarity to g112827663 P18F4.190  
membrane-associated salt-inducible-like protein from  
Arabidopsis thaliana BAC gb|A021637."

gene  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAD30222.1"  
/db\_xref="GI:4835755"

CDS  
/translation="MKPQMFRSYIQFSPSPMORSYSSGNAEFNISGEVSIILAKK  
KPIEPALEVPPLFSKNIITSYIKDEVNROLGREFIIMASRELRBSRESEGLVIDL  
SENGCGLVQMLELKSQSVDSYSCFVYLISAYAMGMAEVAESESFGMKEDCDP  
DVEFYNYILVMMREEVFMFAVYVEMKCNCPULYFEGILMDGLYKRGRTSDR  
KMPDDMGRIISPNRYVYTIILIGICQSGADARLKIFYEMQTSNGYDPSVAHALID  
GCKLGMVAFAELRLFEKQGYVLRGSSILIDGIFRRRTTOAEIYANLKKNI  
KPIILYTIILIGLSKAGKIEDALKLSSMPSKGISPDYICVANAIVALCGRLLEG  
RSLOLEMSETESEPDACTHTILICSMCRNLVRAEETFEIEKSGSPVAFFNALI  
DGCKSGSELKEARLLHKMEVGRPASLELHSHSGNSFPTWESGSLKAYRDLAH  
ADGSSPDIVSYVYLINGFCRAGIDGALKLVNLVOLKSGSPMSTYKTKICLIDETAN  
REERAEFLYAKNDRHSPAVYSISMCRKRVLYAFNLMMKYIKTKICLIDETAN  
ETEOCRKEGETERALRLIELDRKDELITGPYTTILIGICQSGRRFREALMVSYLE  
KILVTPPSCVKLIHGCKREQLDALEVLVLYLDNNFKLMPRCVNLISLLESTEK  
MEIVSQLTNRMERAGYVNDMLREILKVRHROVYLID."

gene  
18900..21756  
/gene="TRK14.5"  
join(18900..18995,19434..19507,19606..19673,19753..19791,  
19928..20028,20168..20280,20369..20468,20547..20693,  
20774..20834,20925..21022,21110..21252,21361..21444,  
21529..21618,21702..21756)  
/note="Is a member of the P100044 glyceraldehyde  
3-phosphate dehydrogenase family. ESTs gb|T43985,  
gb|N38667, gb|N65037, gb|AA713069 and gb|A109548 come  
from this gene."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAD30223.1"  
/db\_xref="GI:4835756"

CDS  
/translation="MAFSSILRSASVAPRDPFSSSDSHSKVLSISGPSRLK  
PSRRSSISSLONGNARSQVPIKATITEVPSAVRRSSSGSKTVGNGRIGRLV  
RIATSRDIEVVANNDPFIADKATWYLYLKTDSHNGKGSINVIDSTLNGKRLV  
VSKRDPSEIEMADIGADYVVESSGVFTLLKASHLGGAKKVIISAPSDAPFVVG  
VNEHTYQPNMDIVSNACITNCLPLKVVHAEESGLEMTYHAATAQKYDGS  
MKMWRGGRGASINIPSSGTAKAVAGVLELNLKLGMAFRVYTSVAVDLCRIE  
KQASVDEYKAIKHAISGPIKGLIGTYDDEVDYSDPFGDSSSIFDANAIGLSKSY  
KLVSWIDNENGISNRVLDLIEHMLVAASH"

gene  
23473..26525  
/gene="TRK14.6"  
join(23473..23665,24252..24364,24454..24699,24777..24857,  
24955..25185,25275..25613,26424..26525)  
/gene="TRK14.6"  
/note="EST gb|AA404917 comes from this gene."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAD30224.1"  
/db\_xref="GI:4835757"

Query Match 14.6% Score 18; DB 8; Length 80374;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gtatgttaatgttagat 21  
|||||  
Db 28318 GTATTGTAAGTGAGAT 28335  
|||||

RESULT 20  
AC008790/c AC008790 117018 bp DNA PRI 09-SEP-2000  
LOCUS Homo sapiens chromosome 5 clone CTD-2040H20, complete sequence.  
DEFINITION AC008790  
ACCESSION AC008790  
VERSION AC008790.6 GI:10044330

gene



KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE 1 (bases 1 to 117018)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 117018)  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
REFERENCE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
TITLE 3 (bases 1 to 117018)  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE Submitted (09-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell  
AUTHORS Drive, Walnut Creek, CA 94598, USA  
TITLE On Sep 9, 2000 this sequence version replaced gi:9256045.  
JOURNAL Draft Sequence Produced by DOE Joint Genome Institute  
COMMENT www.jgi.doe.gov  
www.shc.stanford.edu  
Finishing Completed at Stanford Human Genome Center  
Quality: Phrap Quality >=40 99.8% of Sequence:  
Estimated Total Number of Errors is 0.2.  
Location/Qualifiers  
1. 117018  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2040H20"  
BASE COUNT 34667 a 23973 c 23158 g 35220 t  
ORIGIN

Query Match 14.6%; Score 18; DB 9; Length 117018;  
Best Local Similarity 100.0%; Pred.No.16;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 caacctgatacctaact 45  
|||||  
Db 101045 CAACCTGATGACCTAAT 101028

RESULT 21  
AP003046 140952 bp DNA PLN 08-MAY-2001  
LOCUS AP003046  
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0445D12.  
ACCESSION AP003046  
VERSION AP003046.2 GI:13486797  
KEYWORDS  
SOURCE  
ORGANISM Oryza sativa (cultivar: Nipponbare) DNA, clone:P0445D12.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrharioideae; Oryzaceae; Oryza.  
REFERENCE 1 (sites)  
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.  
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
JOURNAL clone:P0445D12  
2 (bases 1 to 140952)  
Published Only in Database (2000) In press  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Direct Submission  
Submitted (13-DEC-2000) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program, Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
On Mar 28, 2001 this sequence version replaced gi:11862976.  
Genes were predicted from the integrated results of the following:  
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor

(October 1998 version). The genomic sequence was searched against  
NCBI Nonredundant Protein database, nr  
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at  
RGP. Protein homologues of the coding regions were searched against  
NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent  
the identified cDNA sequences using BLASTN 2.0 with the  
corresponding DDBJ accession no. and RGP clone ID.  
A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with EST homology (covering  
almost the entire length of partial sequence) is classified as an  
'unknown' protein. A gene predicted with a gene prediction program  
is classified as a 'hypothetical' protein.  
The orientation of the sequence is from SP6 to T7 of the PAC clone.  
This sequence of P0445D12 clone has an overlap with P0684C01 (DDBJ:  
AP002487) clone at the position 1 to 17,951 of 5' end. The sequence  
of this clone starts at the position 60650 of P0684C01. Detailed  
information on overlap and assembly quality together with  
annotation of this entry is available at  
http://rpg.dna.affrc.go.jp/genomeseq.html.  
Location/Qualifiers  
1. 140952  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="1"  
/clone="P0445D12"  
complement(join(3938..4150,5347..5445,6717..6767))  
/gene="P0445D12.1"  
complement(join(3938..4150,5347..5445,6717..6767))  
/gene="P0445D12.1"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB40030.1"  
/db\_xref="GI:13486798"  
/translation="MTTCSTPLAKGRRRYKHTLVTRAFDCDHLIEHEDDKKQKLY  
ATYKSLAFCLADSYQTCLOKLAQALPNKAMSPKCTVNSPLIITHASQKDRKM  
SKLTWSEKSSRAIIR"  
complement(16002..16439)  
/gene="P0445D12.2"  
complement(16002..16439)  
/gene="P0445D12.2"  
/note="contains ESTs AU093168 (C0321), D15235 (C0321)  
unknown protein"  
/codon\_start=1  
/protein\_id="BAB40031.1"  
/db\_xref="GI:13486799"  
/translation="MSMSMLTSPAEIGRGCPALRKARRSPACAFRRDQYSGAL  
VDSMAVLRREARRKAEENNTETAPAGSAAEKRTYPAIVSDSAAGALQQLMGT  
PSVAIAAALLFAGVPSVAVAHHLAQLAESAIVLQHHVVP"  
complement(join(17738..17923,18232..18316,18417..18629,  
18768..18988,19165..19296,19872..20148,20541..20742,  
21433..21565,21895..22045,22138..22328,22821..23036))  
/gene="P0445D12.3"  
complement(join(17738..17923,18232..18316,18417..18629,  
18768..19165,19296,19872..20148,20541..20742,  
21433..21565,21895..22045,22138..22328,22821..23036))  
/gene="P0445D12.3"  
/codon\_start=1  
/protein\_id="BAB40032.1"  
/db\_xref="GI:13486800"  
/translation="MEVGLGQLLAALAVRAVAGPGPALLPPADDESDAPEA  
GEGGVPPVITRMARITCALKNRGDARILSNAGSEKSGRLALMPSSGSKTT  
LLNVLAGLITSPSLHSGFYINGRPISSEGGYIAVROEDLEFSQUTVRETIICLA  
ELIARLTLPKRRESYVNDLIFRLGVNACDSYGDARVGSISGEEKRSILCEILA  
SPSITFADPEPTTGADARAEKVMETKQLADGRTVYCSIHQPGSYRGFDIVILS  
EGEYVMPARAEPLLYFASLGYCHPCPHVNAEFLADLISVDSASVSSRRLEN  
LIEESNKNVAITESSNLTPNEGSEFSKLQKSTTHRRQWFOPLFRKMQMF  
VDPGNKVRKASVASAIIIFGVSMKQKTSIDQNGMLQVYATWMAALATKTVG  
VPPKRAIVDRERKAGSYALPYSKILAEIPIGAPFPLIFGSLTVPKRIPTPSR  
FAKRGIVTVESPAASMGITVGMAMPTEAAMLAGFSLMTVITVEGCIYVNDNTPV







database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Ch9>  
Rp11-72B4 is from the library RPc1-11.1 constructed by the group of Plieter de Jong. For further details see  
<http://www.chori.org/Bacpac/home.htm>  
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-72B4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
The true left end of clone Rp11-72B4 is at 1 in this sequence. The true left end of clone Rp11-18014 is at 142537 in this sequence. The true right end of clone Rp11-346B7 is at 81734 in this sequence.

## FEATURES

```
source
    Location/Qualifiers
        1..144536
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="9"
        /clone="Rp11-72B4"
        /clone_lib="RPc1-11.1"
        1935..2764
        /note="L2 repeat: matches 1852..2745 of consensus"
        repeat_region
        3353..3419
        /note="Alu repeat: matches 233..298 of consensus"
        repeat_region
        4106..4149
        /note="MIR repeat: matches 103..146 of consensus"
        repeat_region
        5694..5993
        /note="Alusx repeat: matches 1..302 of consensus"
        repeat_region
        6069..6317
        /note="MIR repeat: matches 15..262 of consensus"
        repeat_region
        7021..7128
        /note="MIR repeat: matches 15..262 of consensus"
        repeat_region
        7021..7128
        /note="MIR repeat: matches 15..262 of consensus"
        repeat_region
        7130..7864
        /note="MIR repeat: matches 1..184 of consensus"
        repeat_region
        8159..8304
        /note="L1MC4 repeat: matches 6822..7608 of consensus"
        repeat_region
        8324..8359
        /note="L1MC4 repeat: matches 7813..7952 of consensus"
        repeat_region
        8622..8704
        /note="18 copies 2 mer ac 97% conserved"
        repeat_region
        8622..8704
        /note="MER47 repeat: matches 2242..2323 of consensus"
        repeat_region
        8702..8757
        /note="MER47 repeat: matches 2265..2320 of consensus"
        repeat_region
        9384..9430
        /note="L1M4 repeat: matches 5276..5321 of consensus"
        repeat_region
        9431..9784
        /note="L1M4 repeat: matches 5276..5321 of consensus"
        repeat_region
        9431..9784
        /note="THE1B repeat: matches 1..364 of consensus"
        repeat_region
        9785..9897
        /note="L1M4 repeat: matches 5160..5276 of consensus"
        repeat_region
        9948..10382
        /note="L1M4 repeat: matches 5160..5276 of consensus"
        repeat_region
        9948..10382
        /note="L1M4 repeat: matches 5160..5276 of consensus"
        repeat_region
        10411..10623
        /note="L1M4 repeat: matches 3504..3744 of consensus"
        repeat_region
        10411..10623
        /note="L1M4 repeat: matches 3504..3744 of consensus"
        repeat_region
        10651..10777
        /note="Alub repeat: matches 84..206 of consensus"
        repeat_region
        10794..11175
        /note="Alub repeat: matches 84..206 of consensus"
        repeat_region
        10794..11175
        /note="MIR repeat: matches 1..417 of consensus"
        repeat_region
        11735..12861
        /note="L1M4 repeat: matches 4256..5387 of consensus"
        repeat_region
        11735..12861
        /note="L1M4 repeat: matches 4256..5387 of consensus"
        repeat_region
        12854..13614
        /note="L1M4 repeat: matches 5359..6142 of consensus"
        repeat_region
        14647..15016
        /note="THE1C repeat: matches 1..371 of consensus"
        repeat_region
        15137..15290
        /note="MIR repeat: matches 9..178 of consensus"
        repeat_region
        15137..15290
        /note="MIR repeat: matches 9..178 of consensus"
        repeat_region
        15915..16173
        /note="L1MC2 repeat: matches 5913..6162 of consensus"
        repeat_region
        17298..17542
        /note="MIR repeat: matches 15..262 of consensus"
        repeat_region
        17298..17542
        /note="MIR repeat: matches 15..262 of consensus"
        repeat_region
        17480..17590
        /note="Single clone region. Assembly confirmed by restriction digest data."
        repeat_region
        17729..17848
        /note="60 copies 2 mer tt 60% conserved"
        repeat_region
        17915..18209
        /note="Alu repeat: matches 1..296 of consensus"
        repeat_region
        18257..18527
        /note="Alu repeat: matches 1..296 of consensus"
        repeat_region
        18526..18645
        /note="Alu repeat: matches 5893..6175 of consensus"
        repeat_region
        18526..18645
        /note="Alu repeat: matches 5893..6175 of consensus"
        repeat_region
        18647..18889
        /note="L1 repeat: matches 3985..4104 of consensus"
        repeat_region
        18647..18889
        /note="L1 repeat: matches 3985..4104 of consensus"
        repeat_region
        18915..19315
        /note="Alusx repeat: matches 64..302 of consensus"
        repeat_region
        19316..19616
        /note="L1 repeat: matches 3596..3984 of consensus"
        repeat_region
        19316..19616
        /note="L1 repeat: matches 3596..3984 of consensus"
        repeat_region
        19617..21010
        /note="Alusx repeat: matches 1..301 of consensus"
        repeat_region
        19617..21010
        /note="Alusx repeat: matches 1..301 of consensus"
        repeat_region
        21041..21231
        /note="L1 repeat: matches 2138..3596 of consensus"
        repeat_region
        21041..21231
        /note="L1 repeat: matches 2138..3596 of consensus"
        repeat_region
        21275..21472
        /note="L1 repeat: matches 1910..2104 of consensus"
        repeat_region
        21275..21472
        /note="L1 repeat: matches 1910..2104 of consensus"
        repeat_region
        21473..21774
        /note="L1MC4 repeat: matches 1489..1687 of consensus"
        repeat_region
        21473..21774
        /note="L1MC4 repeat: matches 1489..1687 of consensus"
        repeat_region
        21775..22900
        /note="Alus81 repeat: matches 3..296 of consensus"
        repeat_region
        21775..22900
        /note="Alus81 repeat: matches 3..296 of consensus"
        repeat_region
        22902..23133
        /note="L1MC4 repeat: matches 844..1489 of consensus"
        repeat_region
        22902..23133
        /note="L1MC4 repeat: matches 844..1489 of consensus"
        repeat_region
        23135..24875
        /note="Alusx repeat: matches 1..239 of consensus"
        repeat_region
        23135..24875
        /note="Alusx repeat: matches 1..239 of consensus"
        repeat_region
        24965..25257
        /note="TIGER1 repeat: matches 1..1792 of consensus"
        repeat_region
        24965..25257
        /note="TIGER1 repeat: matches 1..1792 of consensus"
        repeat_region
        25260..25333
        /note="Alu repeat: matches 1..293 of consensus"
        repeat_region
        25260..25333
        /note="Alu repeat: matches 1..293 of consensus"
        repeat_region
        25777..26864
        /note="L1M4 repeat: matches 2510..2574 of consensus"
        repeat_region
        25777..26864
        /note="L1M4 repeat: matches 2510..2574 of consensus"
        repeat_region
        26961..27967
        /note="MER1C repeat: matches 1..1071 of consensus"
        repeat_region
        26961..27967
        /note="MER1C repeat: matches 1..1071 of consensus"
        repeat_region
        27974..28088
        /note="L1M83 repeat: matches 5173..6178 of consensus"
        repeat_region
        27974..28088
        /note="L1M83 repeat: matches 5173..6178 of consensus"
        repeat_region
        30417..30572
        /note="MER5A repeat: matches 46..159 of consensus"
        repeat_region
        30417..30572
        /note="MER5A repeat: matches 46..159 of consensus"
        repeat_region
        30647..30901
        /note="L1MD repeat: matches 21..189 of consensus"
        repeat_region
        30647..30901
        /note="L1MD repeat: matches 21..189 of consensus"
        repeat_region
        30898..31021
        /note="L1M4 repeat: matches 2552..2809 of consensus"
        repeat_region
        30898..31021
        /note="L1M4 repeat: matches 2552..2809 of consensus"
        repeat_region
        31061..32002
        /note="L1M4 repeat: matches 2773..2896 of consensus"
        repeat_region
        31061..32002
        /note="L1M4 repeat: matches 2773..2896 of consensus"
        repeat_region
        32004..32557
        /note="L1MD repeat: matches 179..1438 of consensus"
        repeat_region
        32004..32557
        /note="L1MD repeat: matches 179..1438 of consensus"
        repeat_region
        32601..34100
        /note="L1MC repeat: matches 1761..2326 of consensus"
        repeat_region
        32601..34100
        /note="L1MC repeat: matches 1761..2326 of consensus"
        repeat_region
        34099..34957
        /note="L1MC repeat: matches 2406..3590 of consensus"
        repeat_region
        34099..34957
        /note="L1MC repeat: matches 2406..3590 of consensus"
        repeat_region
        37554..37636
        /note="L1MD1 repeat: matches 5374..6224 of consensus"
        repeat_region
        37554..37636
        /note="L1MD1 repeat: matches 5374..6224 of consensus"
        repeat_region
        37637..37941
        /note="MIR repeat: matches 177..256 of consensus"
        repeat_region
        37637..37941
        /note="MIR repeat: matches 177..256 of consensus"
        repeat_region
        37942..38084
        /note="Alusx repeat: matches 1..304 of consensus"
        repeat_region
        37942..38084
        /note="Alusx repeat: matches 1..304 of consensus"
        repeat_region
        38862..38960
        /note="MIR repeat: matches 22..177 of consensus"
        repeat_region
        38862..38960
        /note="MIR repeat: matches 22..177 of consensus"
        repeat_region
        39151..39429
        /note="L1M4 repeat: matches 4487..4585 of consensus"
        repeat_region
        39151..39429
        /note="L1M4 repeat: matches 4487..4585 of consensus"
        repeat_region
        39570..40180
        /note="Alus repeat: matches 11..294 of consensus"
        repeat_region
        39570..40180
        /note="Alus repeat: matches 11..294 of consensus"
        repeat_region
        40736..40792
        /note="HAL1 repeat: matches 414..1081 of consensus"
        repeat_region
        40736..40792
        /note="HAL1 repeat: matches 414..1081 of consensus"
        repeat_region
        44656..45707
        /note="HAL1 repeat: matches 883..939 of consensus"
        repeat_region
        44656..45707
        /note="HAL1 repeat: matches 883..939 of consensus"
        repeat_region
        50430..50717
        /note="L1MC2 repeat: matches 5193..6224 of consensus"
        repeat_region
        50430..50717
        /note="L1MC2 repeat: matches 5193..6224 of consensus"
        repeat_region
        50912..52140
        /note="AlusC repeat: matches 1..287 of consensus"
        repeat_region
        50912..52140
        /note="AlusC repeat: matches 1..287 of consensus"
        repeat_region
```



repeat\_region /note="TIGER2 repeat: matches 1461. .2718 of consensus"  
52344. .52629  
/note="Alusg repeat: matches 1. .291 of consensus"  
repeat\_region 54520. .54819  
/note="Alusg repeat: matches 1. .298 of consensus"  
repeat\_region 56188. .56301  
/note="L1MC4 repeat: matches 6570. .6692 of consensus"  
repeat\_region 56527. .56852  
/note="L1MB6 repeat: matches 4937. .5258 of consensus"  
repeat\_region 56853. .57142  
/note="Alusg repeat: matches 1. .290 of consensus"  
repeat\_region 57143. .57847  
/note="L1MB6 repeat: matches 5258. .6004 of consensus"  
repeat\_region 57861. .57893  
/note="MER3 repeat: matches 17. .49 of consensus"  
repeat\_region 58211. .58606  
/note="L1MC4 repeat: matches 7031. .7419 of consensus"  
repeat\_region 58550. .58886  
/note="L1MC4 repeat: matches 7261. .7565 of consensus"  
repeat\_region 58890. .59017  
/note="L1MC4 repeat: matches 7699. .7827 of consensus"

Query Match 14.6%; Score 18; DB 9; Length 144536;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 agatgcctacacccgt 35  
|||||  
Db 141822 AGATGCTTACCAACCTG 141805

RESULT 23  
AL359815  
LOCUS AL359815 146571 bp DNA HTG 03-JUL-2001  
DEFINITION Homo sapiens chromosome 1 clone RP11-63N8, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 10 unordered pieces.  
ACCESSION AL359815  
VERSION AL359815.11 GI:14596332  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULFILL.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 146571)  
McLay, K.  
Direct Submission  
Submitted (02-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
Requests: clonerequest@sanger.ac.uk  
On Jul 4, 2001 this sequence version replaced gi:14587003.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: ba63N8  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 6% of reads  
Dye-terminator Big Dye; 93% of reads  
Consensus quality: 143907 bases at least Q40  
Consensus quality: 144683 bases at least Q30  
Consensus quality: 145097 bases at least Q20  
Insert size: 145671; sum-of-contigs  
Insert size: 155726; 1.2% error; agarose-fp  
Quality coverage: 7.35x in Q20 bases; sum-of-contigs Quality  
coverage: 7.63x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 24804: contig of 24804 bp in length  
24805 24904: gap of 100 bp  
24905 35138: contig of 10234 bp in length  
35139 35238: gap of 100 bp  
35239 54888: contig of 19650 bp in length  
54889 54988: gap of 100 bp  
54989 69175: contig of 14187 bp in length  
69176 69275: gap of 100 bp  
69276 84233: contig of 14958 bp in length  
84234 84333: gap of 100 bp  
84334 99092: contig of 14759 bp in length  
99093 99192: gap of 100 bp  
99193 102346: contig of 3154 bp in length  
102347 102446: gap of 100 bp  
102447 116814: contig of 14368 bp in length  
116815 116914: gap of 100 bp  
116915 122988: contig of 6074 bp in length  
122989 123088: gap of 100 bp  
123089 146571: contig of 23483 bp in length.  
Location/Qualifiers  
1. 146571  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-63N8"  
/clone\_1b="RPC1-11.1"  
1. 24804  
/note="assembly\_fragment:01506  
fragment\_chain:1  
clone\_end:77  
vector\_side:left"  
24905. .35138  
/note="assembly\_fragment:01749  
fragment\_chain:1"  
35239. .54888  
/note="assembly\_fragment:01082  
fragment\_chain:1"  
54989. .69175  
/note="assembly\_fragment:02954  
fragment\_chain:1"  
69276. .84233  
/note="assembly\_fragment:01509"  
84334. .99092  
/note="assembly\_fragment:02635  
fragment\_chain:2"  
99193. .102346  
/note="assembly\_fragment:01308  
fragment\_chain:2"  
102447. .116814  
/note="assembly\_fragment:02746  
fragment\_chain:2"  
116915. .122988  
/note="assembly\_fragment:00077  
fragment\_chain:2"  
123089. .146571  
/note="assembly\_fragment:01690  
fragment\_chain:2  
clone\_end:SP6  
vector\_side:right"  
BASE COUNT 34096 a 36702 c 37423 g 37443 t 907 others  
ORIGIN  
Query Match 14.6%; Score 18; DB 2; Length 146571;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 106 aattatccacacaca 123



Db 85629 AATTATTCCACACACA 85646

# RESULT 24

AC034225 148498 bp DNA HTG 19-APR-2001  
 LOCUS Homo sapiens chromosome 5 clone CTD-2124P14, WORKING DRAFT  
 DEFINITION SEQUENCE, 8 ordered pieces.  
 ACCESSION AC034225  
 VERSION AC034225.4 GI:9256727  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 AUTHORS 1 (bases 1 to 148498)  
 TITLE DOE Joint Genome Institute.  
 JOURNAL Sequencing of Human Chromosome 5  
 REFERENCE Unpublished  
 2 (bases 1 to 148498)  
 AUTHORS DOE Joint Genome Institute.  
 JOURNAL Direct Submission  
 Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT On Jul 18, 2000 this sequence version replaced gi:7712076.

-----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov

Project Information  
 Center Project Name: 675539  
 Center clone name: CTRB-H1\_2124P14

-----  
 Summary Statistics  
 Consensus quality: 139048 bases at least Q40  
 Consensus quality: 145965 bases at least Q30  
 Consensus quality: 147145 bases at least Q20  
 Estimated insert size: 150000; pulse-field gel estimation  
 Estimated insert size: 148148; sum-of-contigs estimation  
 Quality coverage: 5.82 in Q20 bases; pulse-field gel estimation  
 Quality coverage: 5.89 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 1 61862: contig of 61862 bp in length  
 61863 61962: gap of unknown length  
 61963 68022: contig of 6060 bp in length  
 68023 68122: gap of unknown length  
 68123 80136: contig of 12014 bp in length  
 80137 80236: gap of unknown length  
 80237 83311: contig of 3075 bp in length  
 83312 83411: gap of unknown length  
 83412 88833: contig of 5422 bp in length  
 88834 88933: gap of unknown length  
 88934 92360: contig of 3427 bp in length  
 92361 92460: gap of unknown length  
 92461 120402: contig of 27942 bp in length  
 120403 120502: gap of unknown length  
 120503 148498: contig of 27996 bp in length.

## FEATURES

source  
 1. 148498  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2124P14"

BASE COUNT 44488 a 30209 c 29485 g 43610 t 706 others  
 ORIGIN

Query Match 14.6%; Score 18; DB 2; Length 148498;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 28 caaccctgacacctaat 45  
 Db 32644 CAACCCGAGACCTTAT 32661

# RESULT 25

AL357252 153405 bp DNA HTG 02-MAY-2001  
 LOCUS Homo sapiens chromosome 1 clone RP11-35C9, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, 10 unordered pieces.

ACCESSION AL357252  
 VERSION AL357252.8 GI:13990058  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 AUTHORS 1 (bases 1 to 153405)  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On May 7, 2001 this sequence version replaced gi:12227371.

## COMMENT

-----Genome Center

Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 Project Information  
 Center project name: BA35C9

-----Summary Statistics  
 Assembly program: XGAP; version 4.5  
 Sequencing vector: plasmid; LB8752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 150971 bases at least Q40  
 Consensus quality: 151464 bases at least Q30  
 Consensus quality: 151889 bases at least Q20  
 Insert size: 152505; sum-of-contigs  
 Insert size: 157709; 4.7% error; agarose-fp  
 Quality coverage: 7.06x in Q20 bases; sum-of-contigs Quality  
 coverage: 6.89x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 21851: contig of 21851 bp in length  
 21852 21951: gap of 100 bp  
 21952 36986: contig of 15035 bp in length  
 36987 37086: gap of 100 bp  
 37087 51405: contig of 14319 bp in length  
 51406 51505: gap of 100 bp  
 51506 54730: contig of 3225 bp in length  
 54731 54830: gap of 100 bp  
 54831 75701: contig of 20871 bp in length  
 75702 75801: gap of 100 bp  
 75802 85440: contig of 9639 bp in length  
 85441 85540: gap of 100 bp  
 85541 116412: contig of 30872 bp in length



```

* 116413 116512: gap of 100 bp
* 116513 141040: contig of 24528 bp in length
* 141041 141140: gap of 100 bp
* 141141 145417: contig of 4277 bp in length
* 145418 145517: gap of 100 bp
* 145518 153405: contig of 7888 bp in length.
FEATURES
    source
        1. 153405
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-35C9"
            /clone_1fb="RP11-11.1"
            1. 21851
                /note="assembly_fragment:00790
                fragment_chain:1
                clone_end:17
                vector_side:left"
            21952. 36986
                /note="assembly_fragment:02366
                fragment_chain:1"
            37087. 51405
                /note="assembly_fragment:02196
                fragment_chain:1"
            51506. 54730
                /note="assembly_fragment:02549
                fragment_chain:1"
            54831. 75701
                /note="assembly_fragment:02618
                fragment_chain:1"
            75802. 85440
                /note="assembly_fragment:01594
                fragment_chain:1"
            85541. 116412
                /note="assembly_fragment:00526
                fragment_chain:2"
            116513. 141040
                /note="assembly_fragment:02336
                fragment_chain:2"
            141141. 145417
                /note="assembly_fragment:02107
                fragment_chain:2"
            145518. 153405
                /note="assembly_fragment:02012
                fragment_chain:2"
BASE COUNT  33074 a 40678 c 42451 g 36301 t 901 others
ORIGIN
Query Match      14.6%; Score 18; DB 2; Length 153405;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 106 aattattccacacaca 123
      |||
Db 37931 AATTATTCACACACA 37948
      |||
RESULT 26
AC023537/c 156795 bp DNA HTG 12-MAR-2000
LOCUS Homo sapiens chromosome 4 clone RP11-394N5 map 4, WORKING DRAFT
DEFINITION AC023537
SEQUENCE AC023537
ACCESSION AC023537.2 GI:7229813
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS SOURCE
ORGANISM Homo sapiens
human.
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156795)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-394N5

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156795)
AUTHORS Anderson,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouknight,B., Brown,A., Burkett,G., Campolongo,A., Castle,A.,
Chopel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domho,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatord,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laigocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McHeeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigglio,J., Vassiliou,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 12, 2000 this sequence version replaced gi:6978216.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I5543
Center clone name: 394.N.5
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 137441 bases at least Q40
Consensus quality: 147789 bases at least Q30
Consensus quality: 151521 bases at least Q20
Insert size: 16700; agarose-fp
Insert size: 154095; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1411: contig of 1411 bp in length
* 1412 1511: gap of 100 bp
* 1512 2957: contig of 1446 bp in length
* 2958 3057: gap of 100 bp
* 3058 5096: contig of 2039 bp in length
* 5097 5196: gap of 100 bp
* 5197 7059: contig of 1863 bp in length
* 7060 7159: gap of 100 bp
* 7160 9463: contig of 2304 bp in length
* 9464 9563: gap of 100 bp
* 9564 11760: contig of 2197 bp in length
* 11761 11860: gap of 100 bp
* 11861 13457: contig of 1597 bp in length
* 13458 13557: gap of 100 bp
* 13558 15559: contig of 2002 bp in length

```



```
* 1560 15659: gap of 100 bp
* 15660 18335: contig of 2676 bp in length
* 18336 18435: gap of 100 bp
* 18436 20122: contig of 1687 bp in length
* 20123 20222: gap of 100 bp
* 20223 22027: contig of 1805 bp in length
* 22028 22127: gap of 100 bp
* 22128 24224: contig of 2097 bp in length
* 24225 24324: gap of 100 bp
* 24325 26969: contig of 2645 bp in length
* 26970 27069: gap of 100 bp
* 27070 30436: contig of 3367 bp in length
* 30437 30536: gap of 100 bp
* 30537 34800: contig of 4264 bp in length
* 34801 34900: gap of 100 bp
* 34901 38851: contig of 3951 bp in length
* 38852 44086: contig of 5135 bp in length
* 44087 44186: gap of 100 bp
* 44187 44893: contig of 5709 bp in length
* 44896 4995: gap of 100 bp
* 4996 57232: contig of 7237 bp in length
* 57233 57332: gap of 100 bp
* 57333 64959: contig of 7627 bp in length
* 64960 65059: gap of 100 bp
* 65060 72868: contig of 7809 bp in length
* 72869 72968: gap of 100 bp
* 72969 81145: contig of 8177 bp in length
* 81146 81245: gap of 100 bp
* 81246 88655: contig of 7410 bp in length
* 88656 88755: gap of 100 bp
* 88756 97086: contig of 8331 bp in length
* 97087 97186: gap of 100 bp
* 97187 106956: contig of 9770 bp in length
* 106957 107056: gap of 100 bp
* 107057 119495: contig of 12439 bp in length
* 119496 119595: gap of 100 bp
* 119596 134366: contig of 14771 bp in length
* 134367 134466: gap of 100 bp
* 134467 156795: contig of 22329 bp in length.
```

## FEATURES

source

```
1. 156795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone.lib="RP11-394N5"
1. 1411
/note="assembly-fragment"
misc_feature 1512..2957
/note="assembly-fragment"
misc_feature 3058..5096
/note="assembly-fragment"
misc_feature 5197..7059
/note="assembly-fragment"
misc_feature 7160..9463
/note="assembly-fragment"
misc_feature 9564..11760
/note="assembly-fragment"
misc_feature 11861..13457
/note="assembly-fragment"
misc_feature 13558..15559
/note="assembly-fragment"
misc_feature 15660..18335
/note="assembly-fragment"
misc_feature 18436..20122
/note="assembly-fragment"
misc_feature 20223..22027
/note="assembly-fragment"
misc_feature 22128..24224
/note="assembly-fragment"
misc_feature 24325..26969
```

```
misc_feature /note="assembly-fragment"
27070..30436
/note="assembly-fragment"
misc_feature 30537..34800
/note="assembly-fragment"
misc_feature 34901..38851
/note="assembly-fragment"
misc_feature clone_end:17
vector_side:left"
misc_feature 38952..44086
/note="assembly-fragment"
misc_feature 44187..4995
/note="assembly-fragment"
misc_feature 4996..57232
/note="assembly-fragment"
misc_feature 57333..64959
/note="assembly-fragment"
misc_feature 65060..72868
/note="assembly-fragment"
misc_feature 72969..81145
/note="assembly-fragment"
misc_feature 81246..88655
/note="assembly-fragment"
misc_feature 88756..97086
/note="assembly-fragment"
misc_feature 97187..106956
/note="assembly-fragment"
misc_feature 107057..119495
/note="assembly-fragment"
misc_feature 119596..134366
/note="assembly-fragment"
clone_end:5p6
vector_side:left"
misc_feature 134467..156795
/note="assembly-fragment"
BASE COUNT 47744 a 27616 c 27899 g 50833 t 2703 others
ORIGIN
```

```
Query Match 14.6% Score 18; DB 2; Length 156795;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 28 caacctgatgacctaat 45
|||||
Db 146844 CAACCTGATGACCTAAT 146827
```

```
RESULT 27
AC024315 165858 bp DNA HTG 26-MAY-2000
LOCUS AC024315/c
DEFINITION Homo sapiens clone RP11-26N3, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
ACCESSION AC024315
VERSION AC024315.3 GI:8072573
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165858)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-26N3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165858)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenesor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
```







TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Day-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Diaper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flag, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunartine, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Hollway, C.,  
Hollins, B., Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J.,  
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W.,  
Louisleged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,  
Ogulu, M., Okwundu, G., Otagunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojupokan, I., Rolfe, M.,  
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshari, N.,  
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,  
Stone, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Washington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 183496)  
Worley, K.C.

Submitted (01-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Oct 2, 2000 this sequence version replaced g1:10047636.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: HMMH  
Center clone name: RP11-333115  
----- Summary Statistics  
Sequencing vector: M13, L08821  
Chemistry: Dye-terminator Bodipy: 61% of reads  
Chemistry: Dye-terminator Big Dye: 39% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 165672 bases at least Q40  
Consensus quality: 174319 bases at least Q30  
Consensus quality: 178820 bases at least Q20  
Estimated insert size: 176791: sum-of-coverage estimation  
Quality coverage: 3.8x in Q20 bases: sum-of-coverage estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 29040: contig of 29040 bp in length  
\* 29041 29140: gap of unknown length  
\* 29141 48361: contig of 19221 bp in length

48362 48461: gap of unknown length  
\* 48462 65324: contig of 17063 bp in length  
\* 65325 65624: gap of unknown length  
\* 65625 82557: contig of 16933 bp in length  
\* 82558 82657: gap of unknown length  
\* 82658 99864: contig of 17207 bp in length  
\* 99865 99964: gap of unknown length  
\* 99965 112153: contig of 12189 bp in length  
\* 112154 112253: gap of unknown length  
\* 112254 123333: contig of 11080 bp in length  
\* 123334 123433: gap of unknown length  
\* 123434 133305: contig of 9872 bp in length  
\* 133306 133405: gap of unknown length  
\* 133406 139739: contig of 6334 bp in length  
\* 139740 139839: gap of unknown length  
\* 139840 147455: contig of 7616 bp in length  
\* 147456 147555: gap of unknown length  
\* 147556 152212: contig of 4657 bp in length  
\* 152213 152312: gap of unknown length  
\* 152313 157750: contig of 5438 bp in length  
\* 157751 157850: gap of unknown length  
\* 157851 161821: contig of 3971 bp in length  
\* 161822 161921: gap of unknown length  
\* 161922 165295: contig of 3374 bp in length  
\* 165296 165395: gap of unknown length  
\* 165396 168585: contig of 3190 bp in length  
\* 168586 168685: gap of unknown length  
\* 168686 171193: contig of 2508 bp in length  
\* 171194 171293: gap of unknown length  
\* 171294 173564: contig of 2271 bp in length  
\* 173565 173664: gap of unknown length  
\* 173665 176006: contig of 2342 bp in length  
\* 176007 176106: gap of unknown length  
\* 176107 177749: contig of 1643 bp in length  
\* 177750 177849: gap of unknown length  
\* 177850 180432: contig of 2583 bp in length  
\* 180433 180532: gap of unknown length  
\* 180533 182100: contig of 1568 bp in length  
\* 182101 182200: gap of unknown length  
\* 182201 183496: contig of 1296 bp in length.

FEATURES  
source  
1. 183496  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-333115"

BASE COUNT 51162 a 39432 c 39724 g 51055 t 2123 others

ORIGIN

Query Match 14.6%; Score 18; DB 2; Length 183496;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 aattatccacacaca 123  
Db 180935 AATTATTCACACACA 180952

RESULT 29  
AC068581  
LOCUS AC068581 190663 bp DNA HTG 23-SEP-2000  
DEFINITION Homo sapiens chromosome 4 clone RP11-598012 map 4, WORKING DRAFT  
SEQUENCE, 22 unordered pieces.  
ACCESSION AC068581  
VERSION AC068581.2 GI:10280770  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 190663)  
Britten, B., Linton, L., Nusbaum, C. and Lander, E.



TITLE Homo sapiens chromosome 4, clone RP11-598012  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 190663)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F., Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Campello,A., Castle,A., Chaoel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N., McCarthy,M., McKean,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,Y., Morrow,J., Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testa,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Sep 23, 2000 this sequence version replaced g1:7705127. All repeats were identified using RepeatMasker:  
Smt,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

## Project Information

Center project name: LI0318

Center clone name: 598\_O\_12

## Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.96071  
Consensus quality: 177632 bases at least Q40  
Consensus quality: 184698 bases at least Q20  
Consensus quality: 186982 bases at least Q20  
Insert size: 194000; agarose-fp  
Insert size: 188563; sum-of-contigs  
Quality coverage: 4.0 in Q20 bases; agarose-fp  
Quality coverage: 4.1 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

* 1 4553: contig of 4553 bp in length
* 4554 4653: gap of 100 bp
* 4654 5713: contig of 1060 bp in length
* 5714 5813: gap of 100 bp
* 5814 9313: contig of 3500 bp in length
* 9314 9413: gap of 100 bp
* 9414 12245: contig of 2832 bp in length
* 12246 12345: gap of 100 bp
* 12346 15186: contig of 2841 bp in length
* 15187 15286: gap of 100 bp
* 15287 19162: contig of 3876 bp in length
* 19163 19262: gap of 100 bp
* 19263 23897: contig of 4635 bp in length
* 23898 23997: gap of 100 bp

```

FEATURES  
source

```

* 23998 27485: contig of 3488 bp in length
* 27486 27585: gap of 100 bp
* 27586 31745: contig of 4160 bp in length
* 31746 31845: gap of 100 bp
* 31846 36461: contig of 4616 bp in length
* 36462 36561: gap of 100 bp
* 36562 42979: contig of 6418 bp in length
* 42980 43079: gap of 100 bp
* 43080 50663: contig of 7584 bp in length
* 50664 50763: gap of 100 bp
* 50764 58952: contig of 8189 bp in length
* 58953 59052: gap of 100 bp
* 59053 67775: contig of 8723 bp in length
* 67776 67875: gap of 100 bp
* 67876 77789: contig of 9914 bp in length
* 77790 77889: gap of 100 bp
* 77890 86920: contig of 9031 bp in length
* 86921 87020: gap of 100 bp
* 87021 110977: contig of 23957 bp in length
* 110978 111077: gap of 100 bp
* 111078 125542: contig of 14465 bp in length
* 125543 125642: gap of 100 bp
* 125643 141160: contig of 15518 bp in length
* 141161 141260: gap of 100 bp
* 141261 154817: contig of 13557 bp in length
* 154818 154917: gap of 100 bp
* 154918 176781: contig of 21864 bp in length
* 176782 176881: gap of 100 bp
* 176882 190663: contig of 13782 bp in length.
Location/Qualifiers
1. .190663
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-598012"
/clone_lib="RPC1-11 Human Male BAC"
1. 4553
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
4654. .5713
/note="assembly-fragment"
5814. 9313
/note="assembly-fragment"
9414. 12245
/note="assembly-fragment"
12346. .15186
/note="assembly-fragment"
15287. .19162
/note="assembly-fragment"
19263. .23897
/note="assembly-fragment"
23998. .27485
/note="assembly-fragment"
27586. .31745
/note="assembly-fragment"
31846. .36461
/note="assembly-fragment"
36562. .42979
/note="assembly-fragment"
43080. .50663
/note="assembly-fragment"
50764. .58952
/note="assembly-fragment"
59053. .67775
/note="assembly-fragment"
67876. .77789
/note="assembly-fragment"
77890. .86920
/note="assembly-fragment"
87021. .110977
/note="assembly-fragment"

```











```

* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 23321: contig of 23321 bp in length
23322 23421: gap of 100 bp
23422 24529: contig of 1108 bp in length
24530 24629: gap of 100 bp
24630 26959: contig of 2330 bp in length
26960 27059: gap of 100 bp
27060 28240: contig of 1181 bp in length
28241 28340: gap of 100 bp
28341 30420: contig of 2080 bp in length
30421 30520: gap of 100 bp
30521 32131: contig of 1611 bp in length
32132 32231: gap of 100 bp
32232 34170: contig of 1939 bp in length
34171 34270: gap of 100 bp
34271 36716: contig of 2446 bp in length
36717 36816: gap of 100 bp
36817 40349: contig of 3533 bp in length
40350 40449: gap of 100 bp
40450 43365: contig of 2916 bp in length
43366 46874: contig of 3409 bp in length
46875 46974: gap of 100 bp
46975 51691: contig of 4717 bp in length
51692 51791: gap of 100 bp
51792 118666: contig of 66875 bp in length
118667 118766: gap of 100 bp
118767 133750: contig of 14984 bp in length
133751 133850: gap of 100 bp
133851 151236: contig of 17386 bp in length
151237 151336: gap of 100 bp
151337 174475: contig of 23139 bp in length
174476 174575: gap of 100 bp
174576 208697: contig of 34122 bp in length
208698 208797: gap of 100 bp
208798 218073: contig of 9276 bp in length.
Location/Qualifiers
1. 218073
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="10"
/map="10"
/clone="RP23-129023"
/clone_lib="RPCI-23 Female Mouse BAC"
1. 23321
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
misc_feature
23422..24529
/note="assembly_fragment"
24630..26959
/note="assembly_fragment"
27060..28240
/note="assembly_fragment"
28341..30420
/note="assembly_fragment"
30521..32131
/note="assembly_fragment"
32232..34170
/note="assembly_fragment"
34271..36716
/note="assembly_fragment"
36817..40349
/note="assembly_fragment"
40450..43365
/note="assembly_fragment"
43466..46874
/note="assembly_fragment"

```

```

misc_feature /note="assembly_fragment"
46975..51691
/note="assembly_fragment"
misc_feature 51792..118666
/note="assembly_fragment"
misc_feature 118767..133750
/note="assembly_fragment"
misc_feature 133851..151236
/note="assembly_fragment"
misc_feature 151337..174475
/note="assembly_fragment"
misc_feature 174576..208697
/note="assembly_fragment"
misc_feature 208798..218073
/note="assembly_fragment"
vector_side:right"
clone_end:T7
BASE COUNT 61957 a 47778 c 47469 g 59125 t 1744 others
ORIGIN

```

```

Query Match 14.6%; Score 18; DB 2; Length 218073;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 74 taagttcacatgtaagg 91
Db 133435 TAACTTTCACATGTAAAG 133452

```

```

RESULT 33
AC079583/c
LOCUS AC079583 226060 bp DNA HTG 02-SEP-2000
DEFINITION Mus musculus clone RP23-961, WORKING DRAFT SEQUENCE, 19 unordered
pieces.
ACCESSION AC079583
VERSION AC079583.1 GI:9964948
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 226060)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 226060)
DOE Joint Genome Institute.
Direct Submission
Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1740473
Center clone name: RPCI-23_961
-----
Summary Statistics
Consensus quality: 209212 bases at least Q40
Consensus quality: 215597 bases at least Q30
Consensus quality: 216943 bases at least Q20
Estimated insert size: 212000; agarose-1p estimation
Estimated insert size: 224260; sum-of-contigs estimation
Quality coverage: 10.49 in Q20 bases; agarose-1p estimation
Quality coverage: 9.92 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```



\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 1158: contig of 1158 bp in length
* 1159 1258: gap of unknown length
* 1259 2666: contig of 1408 bp in length
* 2667 2766: gap of unknown length
* 2767 3796: contig of 1030 bp in length
* 3797 3896: gap of unknown length
* 3897 5394: contig of 1498 bp in length
* 5395 5494: gap of unknown length
* 5495 10678: contig of 5184 bp in length
* 10679 10778: gap of unknown length
* 10779 14611: contig of 3833 bp in length
* 14612 14711: gap of unknown length
* 14712 19148: contig of 4437 bp in length
* 19149 19248: gap of unknown length
* 19249 23203: contig of 3955 bp in length
* 23204 23303: gap of unknown length
* 23304 29122: contig of 5819 bp in length
* 29123 34265: gap of unknown length
* 34266 34365: gap of unknown length
* 34366 44947: contig of 10582 bp in length
* 44948 45048: gap of unknown length
* 45049 54194: contig of 9146 bp in length
* 54194 54293: gap of unknown length
* 54294 66596: contig of 12303 bp in length
* 66597 66696: gap of unknown length
* 66697 77073: contig of 10377 bp in length
* 77074 77173: gap of unknown length
* 77174 97771: contig of 20598 bp in length
* 97772 97871: gap of unknown length
* 97872 116093: contig of 18222 bp in length
* 116094 116193: gap of unknown length
* 116194 141339: contig of 25046 bp in length
* 141340 177145: contig of 35806 bp in length
* 177146 177245: gap of unknown length
* 177246 226060: contig of 48815 bp in length.

```

## FEATURES

```

source
1..226060
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="R23-961"
/clone_lib="R23-961"
BASE COUNT 56590 a 53883 c 54678 g 59093 t 1816 others
ORIGIN

```

Query Match 14.6%; Score 18; DB 2; Length 226060;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 gagggctgactgactg 73  
|||||  
Db 174925 GAGGCTGCACGTGCTG 174908

## RESULT 34

AX044029 349980 bp DNA PAT 24-NOV-2000  
DEFINITION Sequence 108 from Patent WO0066791.  
ACCESSION AX044029  
VERSION AX044029.1 GI:11342913

KEYWORDS  
SOURCE  
ORGANISM

Neisseria meningitidis.  
Neisseria meningitidis  
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
Neisseria.

REFERENCE  
1 (bases 1 to 349980)  
Pizza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J.C.,  
Masiugnani, V., Galeotti, C., Mora, M., Ratti, G., Scarselli, M.,

Scarlato, V., Rappunli, R., Frazer, C.M. and Grandi, G.  
Neisseria genomic sequences and methods of their use  
Patent: WO 0066791-A 108 09-NOV-2000;  
CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES  
source  
1..349980  
/organism="Neisseria meningitidis"  
/db\_xref="taxon:487"

```

/notes="sequence too long, cut in 8 pieces--seq 1: 1 to
349980 349980 bases--seq 108: 300001 to 649980 349980
bases--seq 109: 600001 to 949980 349980 bases--seq 110:
900001 to 1249980 349980 bases--seq 111: 1200001 to 1549980
349980 bases--seq 112: 1500001 to 1849980 349980 bases--seq
113: 1800001 to 2149980 349980 bases--seq 114: 2100001 to
2272325 172325 bases"

```

BASE COUNT 84410 a 84863 c 94187 g 86520 t  
ORIGIN

Query Match 14.6%; Score 18; DB 6; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 ctatctcaattatc 114  
|||||  
Db 103322 CTATCTCAATTATTC 103339

## RESULT 35

AX094499 518 bp DNA PAT 30-MAR-2001  
LOCUS AX094499/c  
DEFINITION Sequence 623 from Patent WO0118547.  
ACCESSION AX094499  
VERSION AX094499.1 GI:13510717

KEYWORDS  
SOURCE

fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;  
Muscophora; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
1 (bases 1 to 518)  
Davies, R.W., Kaiser, K.U. and Yang, M.Y.  
Essential genes and assays relating thereto  
Patent: WO 0118547-A 623 15-MAR-2001;  
THE UNIVERSITY COURT OF THE UNIVERSITY OF GLASGOW (GB)

FEATURES  
source  
1..518  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

BASE COUNT 134 a 127 c 117 g 140 t  
ORIGIN

Query Match 13.8%; Score 17; DB 6; Length 518;  
Best Local Similarity 100.0%; Pred. No. 1; le+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 caatgcgagggctgcac 66  
|||||  
Db 208 CAATGCGAGGGCTGCAC 192

## RESULT 36

AY004261/c  
ID AY004261 standard; DNA; ORG: 2195 BP.

AY004261;  
XX  
XX  
SV AY004261.1

DT 01-NOV-2000 (Rel. 65, Created)  
DT 01-NOV-2000 (Rel. 65, Last updated, Version 1)  
XX







```

gene
/clone="Y58G8A"
2379. .5358
/gene="Y58G8A.2"
Join(2379. .2443,2939. .3603,3779. .5358)
CDS
/gene="Y58G8A.2"
/notes="contains similarity to reverse transcriptases"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF60814.1"
/db_xref="GI:7332127"
/translation="MRKENKLCITRGDRNFEEYQSKYKLIIGDENARVGNKRDENE
RYGTHAMEPRNDTGEILATFCESNRLMTNSQFYIRRRTHVSPDGMFKLEIDNI
LANGKFTVDTSVLPSETNGSDHRLSRNSFNINLAKLOVKKRKPRLKLEATLTA
VSTTIAAPHGDIDEDYKIRVDSIKAAOOSVOPNSNTRSEETNRKRLTA
RSDPNKSLSTICREAVRDEHNEAKVEKHLPQPPRPPIIAELPPIREIATA
INSFPGKAAGSDKITADFLKSGDNVILITGRFKTLESGATIPDMWTSKTLIFK
KQDRELENYRPICLPLVLYKTFKCVLRIRRTLEADPEVQGRSRSPSTIDHIS
VQRLLEGREYQIPLTLVIFDQKAFDSVEPDWDLSQOQVQNGYINILQECYTD
STFTFPHKRVYPIRVGRQGPISPNIFFACLEHVFQQLMKHKGDEREYEGIR
VNGOQLTNLFRAADIVVAHNPRTASOMLEIYEKSSVGLKINTKTVLRNRFTAK
RKVEIKCPMTNTIIDDVNEEYILGRQINDSNLLPELHRRRAAATAFNKSTDO
ITCPRLRANLPDSVTPATYGESEANTFKELAEKVYTHALERVLITLLEQER
NTHREVRKSKLRDPLIHKKKLGMAGHVARIDGRMTTLMDNCPDERKPVGRP
PMRNDSLKREYVTRDAFGQVQLMWSTQA"
10191. .13116
/gene="Y58G8A.1"
Join(10191. .10218,10271. .10440,10953. .11104,11151. .11292,
11346. .11490,11940. .12106,12166. .12319,12367. .12556,
12804. .13116)
/gene="Y58G8A.1"
/notes="contains similarity to Pfam family PF00065
(Neurotransmitter-gated ion-channel), Score=37.5,
E=3.2e-10, N=2; coded for by C. elegans cDNA yk304h6.3;
coded for by C. elegans cDNA yk304h6.5; coded for by C.
elegans cDNA yk468h10.3; coded for by C. elegans cDNA
yk468h10.5"
/codon_start=1
/protein_id="AAF60813.1"
/db_xref="GI:7332126"
/translation="MILRLISIGVILNHFPGNAASQSDERKLEQLAKGVSKRYP
VKVESTVQVAVYLNIAHVEKDEHQTALVGHLMASWTDEYLERKATNISTLSTI
ASSKLMQPALQLYNSARNSWSLYLTGLPAIYVNGKWSRGTFSPFCOPEFTVP
YDOOSCPVITDMVYGLQVNSLSDPATAGYGPITRLISYDDETFNKRHVGMEVD
SMKHCHYMGPKGLDAEPDGNPDWMSLEFGITLKRHLPYFSLTIVMVPSTSLML
LGFMIENFLRVLFLITAIATAGVSNISLMPGSGRTPKRVHIXSILGISAOY
VLIYMSNPLRVLPSHYEPDEGCDITRIPKKSGLTFPTTKGISEPDELNQAQTS
EAEIQEASTIGISNPLAESGSDTDLLNLPNSSEADISERSETLGMPESEKTOES
QKMRRLIEQLAHKRAFFVYFYFLAILVLA"
Complement(20844. .22820)
/gene="Y58G8A.3"
Complement(Join(20844. .20893,20937. .21123,21220. .21426,
22088. .22244,22429. .22620,22807. .22820))
/gene="Y58G8A.3"
/notes="contains similarity of Pfam family PF00001 (7
transmembrane receptor (rhodopsin)), Score=133.4,
E=5.1e-41, N=1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF60815.1"
/db_xref="GI:7332128"
/translation="MSAALDFYIRSIFFLYGLFVGLGIFNGCVLMAVARNKRLQSA
RVNPLNLIFETDLIFVTAIPYPMATMKDAGFSGVMCHLIVPLMSFQVTSMTLT
ATSIDKFLHINDPTKQPSIRQALATPRLITVSTLINPYLMSFHVQSGSPVQVGE
TPTCGHFCDEANMOSNSRKITGTYMLLQFVPMVATITCTFKILQKYSKMDIQA
QPCQSLTQKQSDATSRKKRVNILLIIMAVTFLIGLSFYLFPGKNSQLF"
Complement(23241. .24074)
/gene="Y58G8A.4"
Complement(Join(23241. .23565,23665. .23873,23982. .24074))
/gene="Y58G8A.4"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein Y58G8A.4"
/protein_id="AAF60816.1"

```

```

/clone="Y58G8A"
2379. .5358
/gene="Y58G8A.2"
Join(2379. .2443,2939. .3603,3779. .5358)
CDS
/gene="Y58G8A.2"
/notes="contains similarity to reverse transcriptases"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF60814.1"
/db_xref="GI:7332127"
/translation="MRKENKLCITRGDRNFEEYQSKYKLIIGDENARVGNKRDENE
RYGTHAMEPRNDTGEILATFCESNRLMTNSQFYIRRRTHVSPDGMFKLEIDNI
LANGKFTVDTSVLPSETNGSDHRLSRNSFNINLAKLOVKKRKPRLKLEATLTA
VSTTIAAPHGDIDEDYKIRVDSIKAAOOSVOPNSNTRSEETNRKRLTA
RSDPNKSLSTICREAVRDEHNEAKVEKHLPQPPRPPIIAELPPIREIATA
INSFPGKAAGSDKITADFLKSGDNVILITGRFKTLESGATIPDMWTSKTLIFK
KQDRELENYRPICLPLVLYKTFKCVLRIRRTLEADPEVQGRSRSPSTIDHIS
VQRLLEGREYQIPLTLVIFDQKAFDSVEPDWDLSQOQVQNGYINILQECYTD
STFTFPHKRVYPIRVGRQGPISPNIFFACLEHVFQQLMKHKGDEREYEGIR
VNGOQLTNLFRAADIVVAHNPRTASOMLEIYEKSSVGLKINTKTVLRNRFTAK
RKVEIKCPMTNTIIDDVNEEYILGRQINDSNLLPELHRRRAAATAFNKSTDO
ITCPRLRANLPDSVTPATYGESEANTFKELAEKVYTHALERVLITLLEQER
NTHREVRKSKLRDPLIHKKKLGMAGHVARIDGRMTTLMDNCPDERKPVGRP
PMRNDSLKREYVTRDAFGQVQLMWSTQA"
10191. .13116
/gene="Y58G8A.1"
Join(10191. .10218,10271. .10440,10953. .11104,11151. .11292,
11346. .11490,11940. .12106,12166. .12319,12367. .12556,
12804. .13116)
/gene="Y58G8A.1"
/notes="contains similarity to Pfam family PF00065
(Neurotransmitter-gated ion-channel), Score=37.5,
E=3.2e-10, N=2; coded for by C. elegans cDNA yk304h6.3;
coded for by C. elegans cDNA yk304h6.5; coded for by C.
elegans cDNA yk468h10.3; coded for by C. elegans cDNA
yk468h10.5"
/codon_start=1
/protein_id="AAF60813.1"
/db_xref="GI:7332126"
/translation="MILRLISIGVILNHFPGNAASQSDERKLEQLAKGVSKRYP
VKVESTVQVAVYLNIAHVEKDEHQTALVGHLMASWTDEYLERKATNISTLSTI
ASSKLMQPALQLYNSARNSWSLYLTGLPAIYVNGKWSRGTFSPFCOPEFTVP
YDOOSCPVITDMVYGLQVNSLSDPATAGYGPITRLISYDDETFNKRHVGMEVD
SMKHCHYMGPKGLDAEPDGNPDWMSLEFGITLKRHLPYFSLTIVMVPSTSLML
LGFMIENFLRVLFLITAIATAGVSNISLMPGSGRTPKRVHIXSILGISAOY
VLIYMSNPLRVLPSHYEPDEGCDITRIPKKSGLTFPTTKGISEPDELNQAQTS
EAEIQEASTIGISNPLAESGSDTDLLNLPNSSEADISERSETLGMPESEKTOES
QKMRRLIEQLAHKRAFFVYFYFLAILVLA"
Complement(20844. .22820)
/gene="Y58G8A.3"
Complement(Join(20844. .20893,20937. .21123,21220. .21426,
22088. .22244,22429. .22620,22807. .22820))
/gene="Y58G8A.3"
/notes="contains similarity of Pfam family PF00001 (7
transmembrane receptor (rhodopsin)), Score=133.4,
E=5.1e-41, N=1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF60815.1"
/db_xref="GI:7332128"
/translation="MSAALDFYIRSIFFLYGLFVGLGIFNGCVLMAVARNKRLQSA
RVNPLNLIFETDLIFVTAIPYPMATMKDAGFSGVMCHLIVPLMSFQVTSMTLT
ATSIDKFLHINDPTKQPSIRQALATPRLITVSTLINPYLMSFHVQSGSPVQVGE
TPTCGHFCDEANMOSNSRKITGTYMLLQFVPMVATITCTFKILQKYSKMDIQA
QPCQSLTQKQSDATSRKKRVNILLIIMAVTFLIGLSFYLFPGKNSQLF"
Complement(23241. .24074)
/gene="Y58G8A.4"
Complement(Join(23241. .23565,23665. .23873,23982. .24074))
/gene="Y58G8A.4"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein Y58G8A.4"
/protein_id="AAF60816.1"

Query Match 13.8%; Score 17; DB 3; Length 34544;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 101 tcttcaatttatttcac 117
Db 34189 tcttcaatttatttcac 34205

RESULT 38
AL591603/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jul 19, 2001 this sequence version replaced gi:14329623.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Emr, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-320P15 is from the library RPc1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-320P15. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-100E6 is at 36226 in this sequence.
The true right end of clone RP11-239I6 is at 2000 in this sequence.

FEATURES
Source
1. .36225
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"

```



```

/clone="RP11-320P15"
/clone_lib="RP11-11.2"
repeat_region
/note="LIM4 repeat: matches 4735. .4952 of consensus"
266. .368
/note="LIMC1 repeat: matches 6225. .6318 of consensus"
369. .563
/note="LIMC1 repeat: matches 1. .218 of consensus"
564. .918
/note="RHE1B repeat: matches 1. .364 of consensus"
919. .1073
/note="LIMC1 repeat: matches 218. .365 of consensus"
1076. .1388
/note="LIMC1 repeat: matches 5883. .6227 of consensus"
1390. .1733
/note="LIMC1 repeat: matches 1. .426 of consensus"
1737. .2398
/note="LIMC1 repeat: matches 5219. .5883 of consensus"
2372. .2594
/note="LIM4 repeat: matches 4490. .4731 of consensus"
2628. .4554
/note="LIM4 repeat: matches 2176. .4208 of consensus"
4590. .9234
/note="LIMC2 repeat: matches 1432. .6115 of consensus"
9330. .9451
/note="61 copies 2 mer lt 59% conserved"
9690. .11001
/note="LIMC2 repeat: matches 653. .2052 of consensus"
11318. .11442
/note="LIMC2 repeat: matches 63. .184 of consensus"
11802. .13026
/note="L2 repeat: matches 1035. .2358 of consensus"
13769. .14132
/note="THE1C repeat: matches 3. .359 of consensus"
16645. .16967
/note="LIMC1 repeat: matches 50. .383 of consensus"
18546. .18902
/note="L2 repeat: matches 2365. .2710 of consensus"
19303. .19360
/note="MER1 repeat: matches 1. .63 of consensus"
19987. .20040
/note="MER20 repeat: matches 160. .214 of consensus"
20092. .20224
/note="MIR repeat: matches 95. .221 of consensus"
20332. .20635
/note="LIMC2 repeat: matches 1. .305 of consensus"
20787. .20832
/note="L2 repeat: matches 2665. .2700 of consensus"
21017. .21469
/note="MER8 repeat: matches 8. .557 of consensus"
21997. .22376
/note="MER57B repeat: matches 42. .403 of consensus"
25114. .25468
/note="LIMC2 repeat: matches 1. .374 of consensus"
26081. .26391
/note="LIMC2 repeat: matches 1. .311 of consensus"
26905. .27023
/note="MIR repeat: matches 46. .175 of consensus"
28243. .28715
/note="LIMC2 repeat: matches 5817. .6300 of consensus"
29703. .29873
/note="LIMC2 repeat: matches 91. .271 of consensus"
29991. .30151
/note="MIR repeat: matches 28. .186 of consensus"
30434. .30744
/note="LIMC2 repeat: matches 2. .311 of consensus"
31502. .31823
/note="LIMC2 repeat: matches 4734. .5064 of consensus"
32200. .32235
/note="LIMC2 repeat: matches 5528. .5563 of consensus"
32330. .32461
/note="LIMC2 repeat: matches 3821. .3954 of consensus"
32465. .32859
repeat_region

```

```

/note="LIMC3 repeat: matches 101. .520 of consensus"
32933. .33053
/note="LIMC3 repeat: matches 3692. .3816 of consensus"
33957. .34033
/note="LIMC3 repeat: matches 5464. .5553 of consensus"
34424. .34729
/note="LIMC3 repeat: matches 4696. .5015 of consensus"
35810. .35863
/note="27 copies 2 mer ac 70% conserved"
BASE COUNT 11082 a 6661 c 6907 g 13575 t
ORIGIN
Query Match 13.8%; Score 17; DB 9; Length 38225;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 aaccctgatacctaata 45
Db 1498 AACCTGATGACCTAAT 1482
|||||

```

```

RESULT 39
SPEC16C4/c 44751 bp DNA PLN 10-JUN-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
S.pombe chromosome III cosmid c16c4.
AL031535.1 GI:3560254
60s ribosomal protein L12A: cdc2-cdc8 family serine-threonine
protein kinase; G-beta repeats; gene free region;
glucosamine-6-phosphate isomerase; low complexity gene free region;
n-terminal acetyltransferase; peptidyl-prolyl cis-trans isomerase;
pseudouridylylate synthase; RNA binding protein; rna binding protein;
rp12.1; sol1 family protein; st5; target of the inhibitor
staurosporine; ttf2 type LTR; TPR Domain; transcription factor
subunit; transcription initiation factor ttf1d 60 kd subunit; WD
repeat domain.
fission yeast.

```

```

SOURCE
ORGANISM
Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetaceae; Schizosaccharomycetaceae;
Schizosaccharomycetes.

```

```

REFERENCE
1 (bases 1 to 44751)
Purnelle,B., Goffeau,A., Wood,V., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (03-SEP-1998) European Schizosaccharomycetes genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: Barrell@sanger.ac.uk
and Unite de Biochimie Physiologique, Universite Catholique de
Louvain, Place Croix du Sud 2/20, B-1348 Louvain-la-Neuve, Belgium
Notes:

```

## COMMENT

Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL, <http://www.sanger.ac.uk/projects/S-pombe/>)  
 During 1995 to 1996 about 66% of S. pombe chromosome 1 was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.  
 Protein coding regions (CDS) have been predicted with the help of computer analysis using the GeneMarker program in Pombase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.  
 CDS are numbered using the following system eg SPBC35H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the



sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
cosmid c16c4 is overlapped at the 5' end by cosmid SPC05E4, EM:AL033406.

## FEATURES

source

Location/Qualifiers

1. .44751

/organism="Schizosaccharomyces pombe"

/strain="972h-"

/db\_xref="taxon:4896"

/chromosome="III"

/clone="cosmid c16c4"

/map="III"

1. .426

/gene="SPCC16C4.19"

/note="SPCC5E4.08"

1. .1736

/note="nominal overlap with cosmid SPC05E4, EM:AL033406 S. pombe chromosome 3"

1. .426

/partial

/gene="SPCC16C4.19"

/note="SPCC16C4.19, len:&gt;140, SIMILARITY: Saccharomyces cerevisiae, RMR-YEAST, nase mrp protein component sm1., (198 aa), fasta scores: opt: 142, E(): 0.0035, (26.0%

identity in 146 aa)"

/codon\_start=1

/label="SPCC16C4.19"

/product="hypothetical protein"

/protein\_id="CA871193.1"

/db\_xref="GI:6855454"

/translation="DPKTLHGSCSGSLFIFGRNKLRTISERKARTDLKGEYSR LGNLEINLCVSYHEIPLGVPLRLHOSKOLRAKAIESSTKLESRSKSHNAVAK QQRRLASGLNGLDPRKKDEYAKSTSLSDFMSP"

join(1000. .1307,1391. .1656,1735. .1796,1877. .2581)

/gene="SPCC16C4.01"

join(1000. .1307,1391. .1656,1735. .1796,1877. .2581)

/gene="SPCC16C4.01"

/note="SPCC16C4.01, len:446, SIMILARITY: Saccharomyces cerev isiae, Q003441, hypothetical 49.9 kd protein., (430

aa), fas ta scores: opt: 995, E(): 0. (34.8% identity in

376 aa)"

/codon\_start=1

/label="SPCC16C4.01"

/product="conserved hypothetical protein"

/protein\_id="CA20740.1"

/db\_xref="GI:3560255"

/db\_xref="SPTREMBL:O74446"

/translation="MSNRIGPQSTKTAKLRLLPSTEEFDDFRRODTGREVYSQIPQ IEGSTAKRAEHLGKRHREPLPRVATGCDPRVDLRFPOSRRSSHKTRKODE CTSYSPYNNRETTDLPTLESSRGLNRESSQESLQIFESGLDRNPLFREYFC

FTYGVVLMGTYTIDERRHRLRELGREIEKLEKIDEEVEEFNYITTLTOPRFENFI ALRDASNVMIRLSISHAIAQSVKISLEELVNETIDATKDTQPMIAETRVNLKREI

NMAVGOLFIRININLGQSVLDSPELMETPOLEPIYTAARSYLEINQVALINQVE VIGDLISLMEKQITHTHDESLMEIYVILGLVLLALFSTIVRLADGFLNDRVLYIN

NSNKFLLFAMKLVFYIGLEAVSITTESEFLHYIIIEEMGKKKIKIPYLS"

1308. .1313

/note="1313

/note="gtcgtt, splice donor sequence"

1376. .1390

/note="ctaacattataacag, splice branch and acceptor"

1657. .1662

/note="gtatgc, splice donor sequence"

1715. .1734

/note="tcaacattgtatatttag, splice branch and acceptor"

1797. .1802

/note="gtatga, splice donor sequence"

1864. .1876

/note="ctaacagatatag, splice branch and acceptor"

complement(join(3019. .3212,3265. .3292,3338. .3349))

/gene="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c, len:77"

/gene="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/note="SPCC16C4.20c"

/codon\_start=1  
/label="SPCC16C4.20c  
/product="hypothetical protein"  
/protein\_id="CA871194.1"  
/db\_xref="GI:6855455"

/translation="MDPVVNLQETBEDVYRRKTRIRKRSKSEIEKECBMAVKLYN

VERDIORYRHEAQLSKLEKLNKEKAGE"

complement(3213. .3332)

/note="ctaacattataacag, splice branch and acceptor"

complement(3213. .3264)

/note="confirmed intron"

complement(3259. .3264)

/note="gtcgtt, splice donor sequence"

complement(3293. .3304)

/note="ttgaccgcgtag, splice branch and acceptor"

complement(3332. .3337)

/note="gtatga, splice donor sequence"

complement(4510. .6156)

/gene="SPCC16C4.02c"

complement(4510. .6156)

/gene="SPCC16C4.02c"

/note="SPCC16C4.02c, len:548"

/codon\_start=1

/label="SPCC16C4.02c"

/product="hypothetical protein"

/protein\_id="CA20741.1"

/db\_xref="GI:3560256"

/db\_xref="SPTREMBL:O74447"

/translation="KHIPPHLHGKGYRTISIEGLLEDSDSYASEKISSEHYTEVH

FYVDKEDSNASGSRGSMLENCFSLSLAODTSKPVSLMLALNDPMLIRK

WERMDKFLDRLILSTHYEYVDVGSILAFCSSEALISYEVKRRVSTLQCLKHY

DLCIPVICTLSNPKSAKYLTYTSFINPEPQAFELISNLVYLDNVOYTRPJE

OGIDNRKGLKADCTFESFSDLESFPOVSQSAIRANLOPLMDAVERITDNKLS

AYILSLNKAAGPASIIMPDGFMILIVIRGSRSEIRGSLGMLYKAVGQKRGHTVYT

VECTEVLGLILRYLQENCDLQRIEPPKFRQSLTELESDTDMLFLDADANNK

RDLMSHVTVISAVATLCMLTEDDSQYQASGLMDIFVLMRHSNNGIDAKWLSV

ALPMSLNKVFKAEPDPAKMYVDFTKCNLDLGDGKSFNDYIISTEEDEDERL

AOAIDPHILIOINSLVPOSIMNDIMQEPYKNLLESNF"

join(6836. .7236,7382. .7508)

/gene="SPCC16C4.03"

join(6836. .7236,7382. .7508)

/note="SPCC16C4.03, len:175, SIMILARITY: Emeritella nidulans, Q042735, peptidyl-prolyl cis/trans isomerase., (176 aa), f asta scores: opt: 666, E(): 0. (57.5% identity

in 174 aa)"

/codon\_start=1

/label="SPCC16C4.03"

/product="peptidyl-prolyl cis-trans isomerase."

/protein\_id="CA20742.1"

/db\_xref="GI:3560257"

/db\_xref="SPTREMBL:O74448"

/translation="MSNTGSLPKRWIVKISRNRPFYFNTTHESLMEPPAATDMAAL

KKEFIAELEDSEVTPPEASPKIRASHLIVKHRSRPPSMKEHTITRSSEARKLAE

HKEOILSGSVSHMDLAKMESDSSARRGELGEFRDEMKRFEDAAARALKRGEISG

VETSSGFIILIRHMA"

6851. .6943

/gene="SPCC16C4.03"

/note="Match to PF00397 WW\_rsp5\_WWP, WW domain Score

45.65"

join(7025. .7236,7382. .7505)

/gene="SPCC16C4.03"

/note="Match to PF00639 Rotamase, PPIC-type PPIASE domain.

Score 138.03"

7237. .7242

/note="gtcgtt, splice donor sequence"

7366. .7381

/note="ctaacattatgttag, splice branch and acceptor"

complement(7505. .11633)

/note="possible gene free region, low complexity,

SPCC16C4.14c and SPCC16C4.15c are predicted in this

region, but have very low coding potential and no

homology"







## Query Match

RLAGVTLVAVPVVYVYRNPLPNTSTEGFNLQOKRDQAVRAVHAYAPWLYTMNTQNM  
 PGSSVYNRDGVLSPDKDIIKLGSSRCKPLAEVQOQIHESIINDMIVGFENMED  
 PLEENPFLNREGSVHLMQDDEMLVPVTLQRYIADKLPLHHEVAAGGHHFPLAKG  
 YVDEIVKTLNHH  
 mRNA complement(join<18687..19284,20351..20547,20799..>21119))  
 /gene="F1017.4"  
 complement(18687..21119)  
 /gene="F1017.4"  
 /note="similar to unknown protein GB:AAD21437 from  
 [Arabidopsis thaliana]"  
 CDS complement(join(18687..19284,20351..20547,20799..21119))  
 /gene="F1017.4"  
 /codon\_start=1  
 /product="unknown protein; 21119-18687"  
 /protein\_id="AAG52412.1"  
 /db\_xref="GI:12324910"  
 /translation="MASDSSRSLSLKPPSRKSLFSPSVIVIVGIIATLTYSKLP  
 OPRICSSSGPPTAPRIKLDGRYLAKHGLPREKANRIYFIHSGCCRDADAVA  
 TILSPDLVELGLVYVSPDPCESDPSPRSPLASDIEDLDQSLGSKFYIG  
 KSMGGAAMGCLNKYIPRIAGVTLVAVPVVYVYRNPLPNTSTEGFNLQOKRDQAV  
 RYAHVAPWLYTMNTQNMFGPSSIANRDSLSQSDDIISKGYTRKPRHAEVROGI  
 HESINRDMIVGFENMEDPLDNDPFLNNEGVRHLMQDDEMLVPVTLQRYLAHQLPW  
 VHYEVRSGHFFHFTKGVVDNIVTTLTDTDTIRQ"  
 mRNA complement(join<21829..22626,22877..>23197))  
 /gene="F1017.5"  
 complement(21829..23197)  
 /gene="F1017.5"  
 /note="similar to unknown protein GB:AAD21437 from  
 [Arabidopsis thaliana]"  
 CDS complement(join(21829..22626,22877..23197))  
 /gene="F1017.5"  
 /codon\_start=1  
 /product="unknown protein; 23197-21829"  
 /protein\_id="AAG52399.1"  
 /db\_xref="GI:12324897"  
 /translation="MTRDSSRSLSLKPPSRKSLFSPSVIVIVGIIATLTYSKLP  
 PRLICSSSGPPTAPRIKLDGRYLAKHGLPREKANRIYFIHSGCCRDADAVA  
 TILSPDLVELGLVYVSPDPCESDPSPRSPLASDIEDLDQSLGSKFYIG  
 YSMGGAAMGCLNKYIPRIAGVTLVAVPVVYVYRNPLPNTSTEGFNLQOKRDQAV  
 RYAHVAPWLYTMNTQNMFGPSSIANRDSLSQSDDIISKGYTRKPRHAEVROGI  
 HESINRDMIVGFENMEDPLDNDPFLNNEGVRHLMQDDEMLVPVTLQRYLAHQLPW  
 VHYEVRSGHFFHFTKGVVDNIVTTLTDTDTIRQ"  
 mRNA join(<23489..23520,23750..23785,24176..24404,24499..24554)  
 /gene="F1017.6"  
 23489..24554  
 /gene="F1017.6"  
 /note="similar to ribosomal protein L33B GB:NP\_014877 from  
 [Saccharomyces cerevisiae]"  
 CDS join(23489..23520,23750..23785,24176..24404,24499..24540)  
 /gene="F1017.6"  
 /codon\_start=1  
 /product="putative ribosomal protein; 23489-24540"  
 /protein\_id="AAG52401.1"  
 /db\_xref="GI:12324899"  
 /translation="MKGROGGERVRLYVAGTILGYKRSKSNQVPTSLVQEGVNTQ  
 EVMWYKGRMAYIYKAKTKKNGSHYKCIKGYTRPHGNGVYRAKRTSLNLPKMSMR  
 VRFVWYPSNT"  
 mRNA join(<25509..25911,25995..26563,26663..29047,  
 29144..>29950)  
 /gene="F1017.7"  
 25509..29950  
 /gene="F1017.7"  
 /note="similar to GB:K1AA0361 from [Homo sapiens]"  
 CDS join(25509..25911,25995..26563,26663..29047,29144..29950)  
 /gene="F1017.7"  
 /codon\_start=1  
 /product="putative phosphoribosylformylglycinamide  
 synthase; 25509-29950"  
 /protein\_id="AAG52403.1"  
 /db\_xref="GI:12324901"  
 /translation="MLQKSSWSQMLGWSYRMRTSRLNRTKAVSLRCSAQPKNPKAA

13.8%; Score 17; DB 8; Length 50821;

Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 74 taagttcacatgaaag 90  
 ||||||||||||||||  
 Db 8677 TAAGTTTCACATGTAAG 8661

## RESULT 41

AC018376  
 LOCUS AC018376 59935 bp DNA HTG 13-JUL-2000  
 DEFINITION Homo sapiens clone Rp11-3p5, LOW-PASS SEQUENCE SAMPLING.  
 AC018376  
 VERSION AC018376.2 GI:9121717  
 KEYWORDS HTG; HTGS\_PHASE0.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 59935)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens, clone Rp11-3p5  
 Unpublished  
 2 (bases 1 to 59935)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bede,F.,  
 Boguslavskiy,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M.,  
 Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,  
 Domingo,M., Doyle,M., Fenesior,J., Ferreira,P., Fitzhugh,W.,  
 Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,  
 Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L.,  
 Karatas,A., Klein,J., Landers,T., Lehoczy,J., Lieu,C., Locke,K.,  
 Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,  
 Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,  
 O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R.,  
 Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 Direct Submision  
 Submitted (10-DEC-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 13, 2000 this sequence version replaced gi:6554602.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green,P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
 JOURNAL  
 COMMENT

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L2801  
 Center clone name: 3\_p\_5  
 -----  
 \* NOTE: This record contains 60 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 \* 1 915: contig of 915 bp in length  
 \* \* 916 1015: gap of 100 bp  
 \* \* 1016 1921: contig of 906 bp in length  
 \* \* 1922 2021: gap of 100 bp  
 \* \* 2022 2930: contig of 909 bp in length  
 \* \* 2931 3030: gap of 100 bp  
 \* \* 3031 3907: contig of 877 bp in length



```

* 3908 4007: gap of 100 bp
* 4008 4914: contig of 907 bp in length
* 4915 5014: gap of 100 bp
* 5015 5914: contig of 900 bp in length
* 5915 6014: gap of 100 bp
* 6015 6932: contig of 918 bp in length
* 6933 7032: gap of 100 bp
* 7033 7950: contig of 918 bp in length
* 7951 8050: gap of 100 bp
* 8051 8949: contig of 899 bp in length
* 8950 9049: gap of 100 bp
* 9050 9934: contig of 885 bp in length
* 9935 10034: gap of 100 bp
* 10035 10935: contig of 901 bp in length
* 10936 11035: gap of 100 bp
* 11036 11948: contig of 913 bp in length
* 11949 12048: gap of 100 bp
* 12049 12963: contig of 915 bp in length
* 12964 13063: gap of 100 bp
* 13064 13966: contig of 903 bp in length
* 13967 14066: gap of 100 bp
* 14067 14999: contig of 933 bp in length
* 15000 15099: gap of 100 bp
* 15100 15984: contig of 885 bp in length
* 15985 16084: gap of 100 bp
* 16085 16956: contig of 872 bp in length
* 16957 17056: gap of 100 bp
* 17057 17940: contig of 884 bp in length
* 17941 18040: gap of 100 bp
* 18041 18964: contig of 924 bp in length
* 18965 19064: gap of 100 bp
* 19065 19995: contig of 931 bp in length
* 19996 20095: gap of 100 bp
* 20096 21002: contig of 907 bp in length
* 21003 21102: gap of 100 bp
* 21103 22002: contig of 900 bp in length
* 22003 22102: gap of 100 bp
* 22103 23023: contig of 921 bp in length
* 23024 23123: gap of 100 bp
* 23124 23997: contig of 874 bp in length
* 23998 24097: gap of 100 bp
* 24098 25011: contig of 914 bp in length
* 25012 25111: gap of 100 bp
* 25112 26014: contig of 903 bp in length
* 26015 26114: gap of 100 bp
* 26115 27008: contig of 894 bp in length
* 27009 27108: gap of 100 bp
* 27109 27988: contig of 880 bp in length
* 27989 28088: gap of 100 bp
* 28089 29008: contig of 920 bp in length
* 29009 29108: gap of 100 bp
* 29109 30007: contig of 899 bp in length
* 30008 30107: gap of 100 bp
* 30108 31001: contig of 894 bp in length
* 31002 31101: gap of 100 bp
* 31102 31983: contig of 882 bp in length
* 31984 32083: gap of 100 bp
* 32084 32967: contig of 884 bp in length
* 32968 33067: gap of 100 bp
* 33068 33986: contig of 919 bp in length
* 33987 34086: gap of 100 bp
* 34087 34995: contig of 909 bp in length
* 34996 35095: gap of 100 bp
* 35096 36028: contig of 933 bp in length
* 36029 36128: gap of 100 bp
* 36129 36995: contig of 867 bp in length
* 36996 37095: gap of 100 bp
* 37096 38018: contig of 923 bp in length
* 38019 38118: gap of 100 bp
* 38119 39019: contig of 901 bp in length
* 39020 39119: gap of 100 bp
* 39120 40012: contig of 893 bp in length
* 40013 40112: gap of 100 bp

```

```

* 40113 41016: contig of 904 bp in length
* 41017 41116: gap of 100 bp
* 41117 42005: contig of 889 bp in length
* 42006 42105: gap of 100 bp
* 42106 43055: contig of 950 bp in length
* 43056 43155: gap of 100 bp
* 43156 44088: contig of 933 bp in length
* 44089 44188: gap of 100 bp
* 44189 45107: contig of 919 bp in length
* 45108 45207: gap of 100 bp
* 45208 46128: contig of 921 bp in length
* 46129 46228: gap of 100 bp
* 46229 47142: contig of 914 bp in length
* 47143 47242: gap of 100 bp
* 47243 48115: contig of 873 bp in length
* 48116 48215: gap of 100 bp
* 48216 48936: contig of 721 bp in length
* 48937 49036: gap of 100 bp
* 49037 49922: contig of 886 bp in length
* 49923 50022: gap of 100 bp
* 50023 50946: contig of 924 bp in length
* 50947 51046: gap of 100 bp
* 51047 51950: contig of 904 bp in length
* 51951 52050: gap of 100 bp
* 52051 52962: contig of 912 bp in length
* 52963 53062: gap of 100 bp
* 53063 53922: contig of 860 bp in length
* 53923 54022: gap of 100 bp
* 54023 54959: contig of 937 bp in length
* 54960 55059: gap of 100 bp
* 55060 55924: contig of 865 bp in length
* 55925 56024: gap of 100 bp
* 56025 56942: contig of 918 bp in length
* 56943 57042: gap of 100 bp
* 57043 57936: contig of 894 bp in length
* 57937 58036: gap of 100 bp
* 58037 58933: contig of 897 bp in length
* 58934 59033: gap of 100 bp
* 59034 59935: contig of 902 bp in length.

```

FEATURES  
source 1. 59935  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-3P5"  
/clone\_lib="RPC1-11 Human Male BAC"

BASE COUNT 16503 a 10402 c 9763 g 17088 t 6179 others  
ORIGIN

Query Match 13.8%; Score 17; DB 2; Length 59935;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 aggtatgtcaatgga 18  
DB 5750 AGGTATGTCAATGGA 5766  
|||||

RESULT 42  
AP000381 63604 bp DNA PLN 27-DEC-2000  
LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K17E12.  
DEFINITION AP000381 BA000014  
ACCESSION AP000381.1 GI:5672506  
VERSION  
KEYWORDS  
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:Mitsui TAC  
clone:K17E12.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (sites)  
REFERENCE Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.



TITLE	Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the 4,251,695 bp regions covered by 90 pl. TAC and BAC clones
JOURNAL	DNA Res. 7 (3), 217-221 (2000)
MEDLINE	20363099
REFERENCE	2 (bases 1 to 63604)
AUTHORS	Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
TITLE	Direct Submission
JOURNAL	Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see <a href="http://www.kazusa.or.jp/kaos/cgi-bin/aggd_graph.cgi?c=K17E12">http://www.kazusa.or.jp/kaos/cgi-bin/aggd_graph.cgi?c=K17E12</a> Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <a href="http://compbio.ornl.gov/Grail-1.3/">http://compbio.ornl.gov/Grail-1.3/</a> ), GENSCAN (Chris Burge, MIT, <a href="http://CCR-081.mit.edu/GENSCAN.html">http://CCR-081.mit.edu/GENSCAN.html</a> ), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <a href="http://www.cbs.dtu.dk/services/NetGene2/">http://www.cbs.dtu.dk/services/NetGene2/</a> ) and Slicepredictor (Volker Brendel, Stanford University, <a href="http://gremlini.zool.iastate.edu/cgi-bin/sp.cgi">http://gremlini.zool.iastate.edu/cgi-bin/sp.cgi</a> ). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <a href="http://genome.wustl.edu/eddy/tRNAscan-SE/">http://genome.wustl.edu/eddy/tRNAscan-SE/</a> ). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MYP5 and the 3' clone is K1G2.
FEATURES	Location/Qualifiers
SOURCE	1.63604
CDS	/organism="Arabidopsis thaliana"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/chromosome="3"
	/clone="K17E12"
	/clone_1lb="Mitsui TAC"
	complement(join(3373..3519,3594..3676,3763..3835,3921..4014,4236..4296,4481..4541,4626..4709,4797..4991,5261..5340,5669..5765,5900..6018,6097..6166,6238..6329,6591..6786))
	/note="gene_id:K17E12.1"
	/codon_start=1
	/evidence=not_experimental
CDS	/product="uridine kinase-like protein"
	/protein_id="BAB02114.1"
	/db_xref="GI:9294212"
	/translation="MPEDSTAIIDYVMEKASGPHSGRLDGLTSSPKSSVSSPSHR LNSSFSADVDPAAPHPPIYVGTGASGKTVQCDMTIQLDHRIVLYNOSFIRG NACDVIILEGILVFDHDAFDTQQLHCHVDILKSGQTPQIPITDFTHOKVAFRQ YAKFVRAFDVFLPSKRYADVIIIPRGDNHVAVDLVQHIHTKLGSHDDCKVTPNF VLEETFOIRGMHTLIREKDISKHIDFVSDRLRLVVEHGLHLPTEKQVYPTGS YSGADPCFKIKGVGSVIRSGSEMENARACGKIKIGKILIHEDGDMOLIVKLPD ISERHVILDPVLTGTSANQATEILLOKGVPRAHITFLNLISAPBIRHCVCARFRL KIVTSELDCINLEFRYIRGLGEGDRIFETDE"
	complement(join(7607..7950,8044..8209))
	/note="gene_id:K17E12.2"
	unknown protein"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="BAB02115.1"
CDS	/db_xref="GI:9294213"
	/translation="MKQAVLVILVFSGLTSVKTALAARHVIGSGQWESQVDPDSN SDFSKVQDIELHSVVELSSTAYKSCDGVSNLSGNDVYKLSKTGRYFACGT VGHCEQDMKIKVMVSSDSKSSPSSGSGSGSSSSSHGLRASGCMYFVVGSL VYIGLIMAY"
	complement(join(9327..9715,9806..9886,10692..10784,10950..11091))
	/note="gb AA03438.1"
CDS	gene_id:K17E12.3
	similar to unknown protein"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="BAB02116.1"
	/db_xref="GI:9294214"
	/translation="MGSSSSSLNNSPIRADSMVTPESQMTVNDKNNDVSIPLSVK KSPESPSTSTIPANNLPYVKSWSFSSSKSPGSKDETFPDSOPMLOSDDDPFS VNDFTPSLQNTKSSRSDRPFRHNIIFIEKKPSRSSSSPAPLPRKKIKGLFRDSI REERESGSSSLSTPLSGANSREFMDTIEIEEKKSWMHHRCLPFGSSGSGSM ERRKMSSEPPVAAY"
	complement(join(15611..15775,15859..16065,16142..16247,16334..16501,16582..16811,16889..17074,17245..17463))
	/note="gene_id:K17E12.4"
	unknown protein"
CDS	/codon_start=1
	/evidence=not_experimental
	/protein_id="BAB02117.1"
	/db_xref="GI:9294215"
	/translation="MAKPPHHHHHSSRRMLVLYFTSVIGIFIAFLCISSSIP SVSAVPSIMVPVNRPEIQTPIIDSKITQKSKQSNQDKHVRFLSAIFAIPABELK EEMESAFVRLDGTSVQINMLVFSYSGLDYVHSHVDVFNFDNKCWCRFHPKEM ANSHLGIYDGRVYVVSQGLDQCRCRQTSRSPVLSFTFWLEFSLPAPRAPATQ IWRGLIWMGSKENRNAVAFDHSIAYKDGKALDEMBEEVPIPRGGPHACVANDK LVIYGGGDEPMAKPNSEPIFKSRRRIENGVEYMDDEKMKMLPMPKNNHISA WIIVNNSVIVGCTDHPVYKRLVLYNGELFERPOLDTLTSVIGRLPYRKTMAGFW NGYLIYFSGDRGDPNPQPGKVIQEMRKIKAF"
	18093..19325
	/note="contains similarity to unknown protein"
	gb AA01017.1
	gene_id:K17E12.5"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="BAB02118.1"
	/db_xref="GI:9294216"
CDS	/translation="MGSVSLKIGDGTARFERRTSCSSAVNMLFVYVTTNLFALYAF SSSHQANSPILHSSNNISLVSQHLSTLREIDSSORLAQOMKQCYEIDISRPIT VPEKLEFIORHOLPLGDSRTGITEWYSVGHGCKTDLISQYMSKVPDRCPDMS LGOKILIRACEPLPRRCIAKTYOKODSSPSIMSVSNKSNKNSGGLCKSPDKS GKRLSKECVCFPLGVKDRFVYKKGNDLIDVYLGLSGKTRIGDIGSGSGTFPA RMAEKNTVITNTLNGCAPSEFTIAGGLEPLFLSLDHRPELDNVPDLIHASSGLD YEGRAKLEFLMFLDRLVLRKPRGLMFYCANDEKKKELTRMIERGVKRLKLVIGE KAAQVLYSAVLCKPAA"
	complement(join(20013..20129,20428..20493,20594..20710,20791..20862,20950..21229,21627..21838,22180..22239))
	/note="gene_id:K17E12.6"
	/codon_start=1
	/evidence=not_experimental
	/product="ubiquinol-cytochrome c1 reductase cytochrome c1 precursor-like protein"
	/protein_id="BAB02119.1"
	/db_xref="GI:9294217"
	/translation="MVGSGVYIOILRRKLSQSLAPVLSWFSKKAHEDAGSGVARA LALLGAVTOLLSFTSVASADAEHGLSESEYWPMDGILISVDHASIRGHQVYQOV CASCHSRLISYLDLVQVATLEEKARMAEIVVQGNDEGFTTPRGLSDFRPOP YANESARPSLNGCAYPDPDLITLKRANGNYYFALITGRDPAGISIREGLHYNV FPGGATAMPKMLNDEAVYEDGVPAPEAOQKQIVSLAQAAPBEMERKLMGKTFE LLSLALQAAVYRRLKNSVLSKRLIVDVVN"
	23502..26350
	/note="emb CAB62340.1"
	gene_id:K17E12.7
	similar to unknown protein"
CDS	/codon_start=1
	/evidence=not_experimental
	/protein_id="BAB02120.1"
	/db_xref="GI:9294218"
	/translation="MDQDDWILGLRYRAGKADKVSVDALMLRYRPAKPTGQPCGV ADNNNSYSYMSKRTKRYKRVAKNNKGTGRGSRSDLSDRQDTAVVTLQLPEKSD ISGEVSPILQDLSIDPSYKSLIGETQETNNWGNPNSLTEMGVNVAEVSVCES LSASHVAGITDVELVDNIGKDTCPAFPSDGSNNRYVNEAIRRVASGDDSDASISPEV VMLVAEBATAMHCNTQAFTCRVMQYTKETRYTTKTVPCDVKMEGGGAWRLDPTTA ALTLWL"
	join(32280..33512,33590..33632,33728..33790,34020..34090,



```

34154 . 34241,34515 . 34595,34685 . 34794,35165 . 35234,
35378 . 35520,35718 . 36272)
/notes="gb|AA01563.1
gene_id:k17e12.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02121.1"
/db_xref="GI:9294219"
/translation="MWSAAPFGGYRNTPEAPESESGSSAOIDTEVYASNSSTP
ARKCIMLNSNDDEPGVROVISLYNMSOERDLRYRKLLEQRIYKNEIQRM
NPAVSTSDRYGSESGKISSRVSNSKPSDPAVSGSKRHQNRQNTSGMSTGSK
FESSKETHSTRTNITLMKOCDTILRLKSHPHSVQAPDYVKRLNIPDTITIKPM
DGTIVKRLNLSGVSSPHEFADVRLFTNMYTPNGDHVMGDLISLFEARKMT
IKKLPCCSMQTLPAVTLTPNDEKKAISVPAPKRRKMSPVRESVPEYKPLMTVE
RRLRGLESLDELPAHITIDFLKHNNSGGEIAEDEIIDLVDLDEVYLYRLNLD
EYIIONKEAKOTVEPEEILINGSRNSSLONGMADEYIDVNGPISRSSDSDS
GSSGSDODAKPMVGDSKLVNLOWPFTANSFQDENTRIDDLVGSOSTALOM
DICSQOKSSDSDGHECNILETPASSSEKRYAALTKRPAIILIKARKRPONGI
KQDPERLKERELVLOKKEKARLOAEADARKOAEAEAAAEAKRRELE
REARQALIKMEKYEIINENSRPLEDEMSSAPOLPSAETSPEPRLDLSGN
LRGNSPLEQLGLYMKODDEEPEAPAVKPPDSTERYPLDALGSFKLGCPNLELG
LYMKODDGESEPEAPVVKPNTSLERPVDAGSFNLKGSNPLQGLYMKODDDEE
PEAPAVPLANDVEGEYD"
37184 . 37933
/notes="gene_id:k17e12.9
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02121.1"
/db_xref="GI:9294220"
/translation="MEPTYSKDLPLFESIFTFVYLLGYLFIKKWTPETPLASSCL
ISLHGSAVAVLTATNALSDPNRGSFSSVNTGSONSLDESSAFDLADIVLAFPSA
GSDSLFAAHHLAVLEFVLCGRYVAGACALLALLVVAATSCQNTWTLADRGDA
PLAVSLHREVTVPEFVYASVSGCVLAPLLIKWTFYVSGADVDIPRMVMSWTVI
VTAVVSTIMINIVLFEORVCRKTKRVR"
join(40773..41084,41393..41920)
/notes="gb|AAC49691.1
gene_id:k17e12.10"
/codon_start=1
/evidence=not_experimental

Query Match 13.8%; Score 17; DB 8; Length 63604;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 95 tactatctcaattta 111
|||||
Db 34800 TACTTATCTTCATTTA 34816

RESULT 43
AC018413/C AC018413 65857 bp DNA HTG 12-JAN-2001
LOCUS DEFINITION Homo sapiens clone Rp11-17M16, LOW-PASS SEQUENCE SAMPLING.
AC018413 AC018413 GI:12084068
VERSION HTG; HTGS_PHASE0.
KEYWORDS SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 65857)
AUTHORS Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguslavsky,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
Domino,M., Doyle,M., Fenestor,J., Ferreira,P., Fitzhugh,W.,
Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,

```

```

TITLE
JOURNAL
COMMENT
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L3661
Center Clone name: 17_M16

NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
725 824: contig of 724 bp in length
825 1548: contig of 724 bp in length
1549 1648: gap of 100 bp
1649 2375: contig of 727 bp in length
2376 2475: gap of 100 bp
2476 3224: contig of 743 bp in length
3225 3324: gap of 100 bp
3325 4044: contig of 720 bp in length
4045 4144: gap of 100 bp
4145 4864: contig of 720 bp in length
4865 4964: gap of 100 bp
4965 5687: contig of 723 bp in length
5688 5787: gap of 100 bp
5788 6498: contig of 711 bp in length
6499 6598: gap of 100 bp
6599 7310: contig of 712 bp in length
7311 7410: gap of 100 bp
7411 8149: contig of 739 bp in length
8150 8249: gap of 100 bp
8250 8993: contig of 744 bp in length
8994 9093: gap of 100 bp
9094 9822: contig of 729 bp in length
9823 9922: gap of 100 bp
9923 10654: contig of 732 bp in length
10655 10754: gap of 100 bp
10755 11498: contig of 744 bp in length
11499 11598: gap of 100 bp
11599 12337: contig of 739 bp in length
12338 12437: gap of 100 bp
12438 13134: contig of 697 bp in length
13135 13234: gap of 100 bp
13235 13951: contig of 717 bp in length
13952 14051: gap of 100 bp
14052 14777: contig of 726 bp in length
14778 14877: gap of 100 bp
14878 15594: contig of 717 bp in length

```



```

* 15595 15694: gap of 100 bp
* 15695 16405: contig of 711 bp in length
* 16406 16505: gap of 100 bp
* 16506 17186: contig of 681 bp in length
* 17187 17286: gap of 100 bp
* 17287 18018: contig of 732 bp in length
* 18019 18118: gap of 100 bp
* 18119 18843: contig of 725 bp in length
* 18844 18943: gap of 100 bp
* 18944 19668: contig of 725 bp in length
* 19669 19768: gap of 100 bp
* 19769 20505: contig of 737 bp in length
* 20506 20605: gap of 100 bp
* 20606 21340: contig of 735 bp in length
* 21341 21440: gap of 100 bp
* 21441 22150: contig of 710 bp in length
* 22151 22250: gap of 100 bp
* 22251 22981: contig of 731 bp in length
* 22982 23081: gap of 100 bp
* 23082 23812: contig of 731 bp in length
* 23813 23912: gap of 100 bp
* 23913 24631: contig of 719 bp in length
* 24632 24731: gap of 100 bp
* 24732 25459: contig of 728 bp in length
* 25460 25559: gap of 100 bp
* 25560 26221: contig of 662 bp in length
* 26222 26321: gap of 100 bp
* 26322 27049: contig of 728 bp in length
* 27050 27149: gap of 100 bp
* 27150 27886: contig of 737 bp in length
* 27887 27986: gap of 100 bp
* 27987 28708: contig of 722 bp in length
* 28709 28808: gap of 100 bp
* 28809 29536: contig of 728 bp in length
* 29537 29636: gap of 100 bp
* 29637 30363: contig of 727 bp in length
* 30364 30463: gap of 100 bp
* 30464 31182: contig of 719 bp in length
* 31183 31282: gap of 100 bp
* 31283 31991: contig of 709 bp in length
* 31992 32091: gap of 100 bp
* 32092 32811: contig of 720 bp in length
* 32812 32911: gap of 100 bp
* 32912 33622: contig of 711 bp in length
* 33623 33722: gap of 100 bp
* 33723 34412: contig of 690 bp in length
* 34413 34512: gap of 100 bp
* 34513 35248: contig of 736 bp in length
* 35249 35348: gap of 100 bp
* 35349 36092: contig of 744 bp in length
* 36093 36192: gap of 100 bp
* 36193 36909: contig of 717 bp in length
* 36910 37009: gap of 100 bp
* 37010 37740: contig of 731 bp in length
* 37741 37840: gap of 100 bp
* 37841 38573: contig of 733 bp in length
* 38574 38673: gap of 100 bp
* 38674 39380: contig of 707 bp in length
* 39381 39480: gap of 100 bp
* 39481 40201: contig of 721 bp in length
* 40202 40301: gap of 100 bp
* 40302 41017: contig of 716 bp in length
* 41018 41117: gap of 100 bp
* 41118 41820: contig of 703 bp in length
* 41821 41920: gap of 100 bp
* 41921 42648: contig of 728 bp in length
* 42649 42748: gap of 100 bp
* 42749 43483: contig of 735 bp in length
* 43484 43583: gap of 100 bp
* 43584 44321: contig of 738 bp in length
* 44322 44421: gap of 100 bp
* 44422 45147: contig of 726 bp in length
* 45148 45247: gap of 100 bp

```

```

* 45248 45965: contig of 718 bp in length
* 45966 46065: gap of 100 bp
* 46066 46816: contig of 751 bp in length
* 46817 46916: gap of 100 bp
* 46917 47574: contig of 658 bp in length
* 47575 47674: gap of 100 bp
* 47675 48405: contig of 731 bp in length
* 48406 48505: gap of 100 bp
* 48506 49200: contig of 695 bp in length
* 49201 49300: gap of 100 bp
* 49301 50023: contig of 723 bp in length
* 50024 50123: gap of 100 bp
* 50124 50855: contig of 732 bp in length
* 50856 50955: gap of 100 bp
* 50956 51656: contig of 701 bp in length
* 51657 51756: gap of 100 bp
* 51757 52503: contig of 747 bp in length
* 52504 52603: gap of 100 bp
* 52604 53342: contig of 739 bp in length
* 53343 53442: gap of 100 bp
* 53443 54212: contig of 770 bp in length
* 54213 54312: gap of 100 bp
* 54313 55040: contig of 728 bp in length
* 55041 55140: gap of 100 bp
* 55141 55860: contig of 720 bp in length
* 55861 55960: gap of 100 bp
* 55961 56702: contig of 742 bp in length
* 56703 56802: gap of 100 bp
* 56803 57556: contig of 754 bp in length
* 57557 57656: gap of 100 bp
* 57657 58386: contig of 730 bp in length
* 58387 58486: gap of 100 bp
* 58487 59224: contig of 738 bp in length
* 59225 59324: gap of 100 bp

```

```

Query Match      13.8%; Score 17; DB 2; Length 65857;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3  ggtatgttaattgag 19
Db      4032  GGTATGTGTAATGAG 4016

```

```

RESULT  44
LOCUS   AC023633  65998 bp  DNA  HTG  13-JUL-2000
DEFINITION Homo sapiens clone RP11-287024, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC023633
VERSION   AC023633.2  GI:9144869
KEYWORDS  HTG; HTGS_PHASE0.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 65998)
Birken,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-287024
Unpublished
REFERENCE 2 (bases 1 to 65998)
AUTHORS  Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouknighter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fensterlo,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Margus,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,

```



TITLE  
JOURNAL  
COMMENT

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M.,  
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,  
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,  
Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,  
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
Zody,M.

## Direct Submission

Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6980301.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Genome Center

Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information

Center project name: L5205  
Center clone name: 287\_J\_24

NOTE: This record contains 75 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 769: contig of 769 bp in length  
\* 770 869: gap of 100 bp  
\* 870 1650: contig of 781 bp in length  
\* 1651 1750: gap of 100 bp  
\* 1751 2538: contig of 788 bp in length  
\* 2539 2638: gap of 100 bp  
\* 2639 3437: contig of 799 bp in length  
\* 3438 3537: gap of 100 bp  
\* 3538 4338: contig of 801 bp in length  
\* 4339 4438: gap of 100 bp  
\* 4439 5234: contig of 796 bp in length  
\* 5235 5334: gap of 100 bp  
\* 5335 6108: contig of 774 bp in length  
\* 6109 6208: gap of 100 bp  
\* 6209 6972: contig of 764 bp in length  
\* 6973 7072: gap of 100 bp  
\* 7073 7701: contig of 629 bp in length  
\* 7702 7801: gap of 100 bp  
\* 7802 8604: contig of 803 bp in length  
\* 8605 8704: gap of 100 bp  
\* 8705 9481: contig of 777 bp in length  
\* 9482 9581: gap of 100 bp  
\* 9582 10290: contig of 709 bp in length  
\* 10291 10390: gap of 100 bp  
\* 10391 11174: contig of 784 bp in length  
\* 11175 11274: gap of 100 bp  
\* 11275 12066: contig of 792 bp in length  
\* 12067 12166: gap of 100 bp  
\* 12167 12931: contig of 765 bp in length  
\* 12932 13031: gap of 100 bp  
\* 13032 13781: contig of 750 bp in length  
\* 13782 13881: gap of 100 bp  
\* 13882 14655: contig of 774 bp in length  
\* 14656 14755: gap of 100 bp  
\* 14756 15591: contig of 836 bp in length  
\* 15592 15691: gap of 100 bp  
\* 15692 16473: contig of 782 bp in length  
\* 16474 16573: gap of 100 bp

16574 17358: contig of 785 bp in length  
\* 17359 17458: gap of 100 bp  
\* 17459 18266: contig of 808 bp in length  
\* 18267 18366: gap of 100 bp  
\* 18367 19147: contig of 781 bp in length  
\* 19148 19247: gap of 100 bp  
\* 19248 20033: contig of 786 bp in length  
\* 20034 20133: gap of 100 bp  
\* 20134 20934: contig of 801 bp in length  
\* 20935 21034: gap of 100 bp  
\* 21035 21822: contig of 788 bp in length  
\* 21823 21922: gap of 100 bp  
\* 21923 22720: contig of 798 bp in length  
\* 22721 22820: gap of 100 bp  
\* 22821 23610: contig of 790 bp in length  
\* 23611 23710: gap of 100 bp  
\* 23711 24480: contig of 770 bp in length  
\* 24481 24580: gap of 100 bp  
\* 24581 25378: contig of 796 bp in length  
\* 25379 25478: gap of 100 bp  
\* 25479 26278: contig of 800 bp in length  
\* 26279 26378: gap of 100 bp  
\* 26379 27162: contig of 784 bp in length  
\* 27163 27262: gap of 100 bp  
\* 27263 28041: contig of 779 bp in length  
\* 28042 28141: gap of 100 bp  
\* 28142 28942: contig of 801 bp in length  
\* 28943 29042: gap of 100 bp  
\* 29043 29840: contig of 798 bp in length  
\* 29841 29940: gap of 100 bp  
\* 29941 30733: contig of 793 bp in length  
\* 30734 30833: gap of 100 bp  
\* 30834 31632: contig of 799 bp in length  
\* 31633 31732: gap of 100 bp  
\* 31733 32530: contig of 796 bp in length  
\* 32531 32630: gap of 100 bp  
\* 32631 33422: contig of 792 bp in length  
\* 33423 33522: gap of 100 bp  
\* 33523 34280: contig of 758 bp in length  
\* 34281 34380: gap of 100 bp  
\* 34381 35184: contig of 804 bp in length  
\* 35185 35284: gap of 100 bp  
\* 35285 36073: contig of 788 bp in length  
\* 36074 36173: gap of 100 bp  
\* 36174 36967: contig of 794 bp in length  
\* 36968 37067: gap of 100 bp  
\* 37068 37838: contig of 771 bp in length  
\* 37839 37938: gap of 100 bp  
\* 37939 38703: contig of 765 bp in length  
\* 38704 38803: gap of 100 bp  
\* 38804 39549: contig of 746 bp in length  
\* 39550 39649: gap of 100 bp  
\* 39650 40409: contig of 760 bp in length  
\* 40410 40509: gap of 100 bp  
\* 40510 41284: contig of 775 bp in length  
\* 41285 41384: gap of 100 bp  
\* 41385 42161: contig of 777 bp in length  
\* 42162 42261: gap of 100 bp  
\* 42262 43048: contig of 787 bp in length  
\* 43049 43148: gap of 100 bp  
\* 43149 43910: contig of 762 bp in length  
\* 43911 44010: gap of 100 bp  
\* 44011 44774: contig of 764 bp in length  
\* 44775 44874: gap of 100 bp  
\* 44875 45644: contig of 770 bp in length  
\* 45645 45744: gap of 100 bp  
\* 45745 46598: contig of 854 bp in length  
\* 46599 46698: gap of 100 bp  
\* 46699 47466: contig of 768 bp in length  
\* 47467 47566: gap of 100 bp  
\* 47567 48338: contig of 772 bp in length  
\* 48339 48438: gap of 100 bp  
\* 48439 49230: contig of 792 bp in length







\* 20370 20469: gap of 100 bp  
\* 20470 21294: contig of 825 bp in length  
\* 21295 21394: gap of 100 bp  
\* 21395 22195: contig of 801 bp in length  
\* 22196 22295: gap of 100 bp  
\* 22296 23026: contig of 731 bp in length  
\* 23027 23126: gap of 100 bp  
\* 23127 23864: contig of 738 bp in length  
\* 23865 23964: gap of 100 bp  
\* 23965 24741: contig of 777 bp in length  
\* 24742 24841: gap of 100 bp  
\* 24842 25622: contig of 781 bp in length  
\* 25623 25722: gap of 100 bp  
\* 25723 26497: contig of 775 bp in length  
\* 26498 26597: gap of 100 bp  
\* 26598 27257: contig of 660 bp in length  
\* 27258 27357: gap of 100 bp  
\* 27358 28156: contig of 799 bp in length  
\* 28157 28256: gap of 100 bp  
\* 28257 29034: contig of 778 bp in length  
\* 29035 29134: gap of 100 bp  
\* 29135 29924: contig of 790 bp in length  
\* 29925 30024: gap of 100 bp  
\* 30025 30801: contig of 777 bp in length  
\* 30802 30901: gap of 100 bp  
\* 30902 31673: contig of 772 bp in length  
\* 31674 31773: gap of 100 bp  
\* 31774 32558: contig of 785 bp in length  
\* 32559 32658: gap of 100 bp  
\* 32659 33431: contig of 773 bp in length  
\* 33432 33531: gap of 100 bp  
\* 33532 34318: contig of 787 bp in length  
\* 34319 34418: gap of 100 bp  
\* 34419 35198: contig of 780 bp in length  
\* 35199 35298: gap of 100 bp  
\* 35299 36068: contig of 770 bp in length  
\* 36069 36168: gap of 100 bp  
\* 36169 36970: contig of 802 bp in length  
\* 36971 37070: gap of 100 bp  
\* 37071 37861: contig of 791 bp in length  
\* 37862 37961: gap of 100 bp  
\* 37962 38739: contig of 778 bp in length  
\* 38740 38839: gap of 100 bp  
\* 38840 39633: contig of 794 bp in length  
\* 39634 39733: gap of 100 bp  
\* 39734 40532: contig of 799 bp in length  
\* 40533 40632: gap of 100 bp  
\* 40633 41403: contig of 771 bp in length  
\* 41404 41503: gap of 100 bp  
\* 41504 42297: contig of 794 bp in length  
\* 42298 42397: gap of 100 bp  
\* 42398 43186: contig of 789 bp in length  
\* 43187 43286: gap of 100 bp  
\* 43287 44087: contig of 801 bp in length  
\* 44088 44187: gap of 100 bp  
\* 44188 44929: contig of 742 bp in length  
\* 44930 45029: gap of 100 bp  
\* 45030 45844: contig of 815 bp in length  
\* 45845 45944: gap of 100 bp  
\* 45945 46731: contig of 787 bp in length  
\* 46732 46831: gap of 100 bp  
\* 46832 47632: contig of 801 bp in length  
\* 47633 47732: gap of 100 bp  
\* 47733 48515: contig of 783 bp in length  
\* 48516 48615: gap of 100 bp  
\* 48616 49377: contig of 762 bp in length  
\* 49378 49477: gap of 100 bp  
\* 49478 50263: contig of 786 bp in length  
\* 50264 50363: gap of 100 bp  
\* 50364 51145: contig of 782 bp in length  
\* 51146 51245: gap of 100 bp  
\* 51246 52018: contig of 773 bp in length  
\* 52019 52118: gap of 100 bp

\* 52119 52915: contig of 797 bp in length  
\* 52916 53015: gap of 100 bp  
\* 53016 53823: contig of 808 bp in length  
\* 53824 53923: gap of 100 bp  
\* 53924 54700: contig of 777 bp in length  
\* 54701 54800: gap of 100 bp  
\* 54801 55599: contig of 799 bp in length  
\* 55600 55699: gap of 100 bp  
\* 55700 56493: contig of 794 bp in length  
\* 56494 56593: gap of 100 bp  
\* 56594 57385: contig of 792 bp in length  
\* 57386 57485: gap of 100 bp  
\* 57486 58276: contig of 791 bp in length  
\* 58277 58376: gap of 100 bp  
\* 58377 59163: contig of 787 bp in length  
\* 59164 59263: gap of 100 bp  
\* 59264 60049: contig of 786 bp in length  
\* 60050 60149: gap of 100 bp  
\* 60150 60921: contig of 772 bp in length  
\* 60922 61021: gap of 100 bp

Query Match 13.8%; Score 17; DB 2; Length 70677;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 107 attatccacacaca 123  
|||||  
DB 32737 ATTATTCACACACACA 32721

Search completed: March 19, 2002, 11:05:47  
Job time: 4021 sec







**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 10:18:46 ; Search time 101.49 Seconds

(without alignments)  
1039.029 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123  
Sequence: 1 caggtatgttaatgtgtgaga.....tcaattattccacacaca 123

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq-1101:\*

- 1: /SIDS2/gcgcdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /SIDS2/gcgcdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /SIDS2/gcgcdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /SIDS2/gcgcdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /SIDS2/gcgcdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /SIDS2/gcgcdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /SIDS2/gcgcdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /SIDS2/gcgcdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /SIDS2/gcgcdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /SIDS2/gcgcdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /SIDS2/gcgcdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /SIDS2/gcgcdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /SIDS2/gcgcdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /SIDS2/gcgcdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /SIDS2/gcgcdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /SIDS2/gcgcdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /SIDS2/gcgcdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /SIDS2/gcgcdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /SIDS2/gcgcdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /SIDS2/gcgcdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /SIDS2/gcgcdata/geneseq/geneseq/NA2000.DAT:\*
- 22: /SIDS2/gcgcdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	14.6	464	21	Human colon cancer
2	18	14.6	1303	22	C albicans apoptos
3	18	14.6	9941	21	N. meningitidis pa
4	18	14.6	349980	21	Neisseria meningit
5	18	14.6	1437668	21	N. meningitidis B
6	17	13.8	518	22	Drosophila melanog
7	16	13.0	446	22	Probe #3879 used t
8	16	13.0	459	22	Probe #4282 for ge
9	16	13.0	459	22	Probe #4409 used t
10	16	13.0	459	22	Probe #4170 used t
11	16	13.0	852	20	DNA encoding OMP-1

C 12	16	13.0	1278	21	AAA38323	Human angiotensino
C 13	16	13.0	1278	21	AAA38324	Human angiotensino
C 14	16	13.0	1491	14	AAQ47075	21BA/rthropxy anti
C 15	16	13.0	2732	21	AAQ98971	Human pancreatic c
C 16	16	13.0	4201	21	AAQ78064	Human cancer assoc
C 17	16	13.0	4201	22	AAH33277	Human angiotensino
C 18	16	13.0	5308	22	AAQ91600	Human angiotensino
C 19	16	13.0	5308	22	AAQ91670	Human angiotensino
C 20	16	13.0	5308	22	AAQ91671	Human angiotensino
C 21	16	13.0	5308	22	AAQ91672	Human angiotensino
C 22	16	13.0	5308	22	AAQ91674	Human angiotensino
C 23	16	13.0	5308	22	AAQ91675	Human angiotensino
C 24	16	13.0	5308	22	AAQ91676	Human angiotensino
C 25	16	13.0	5308	22	AAQ91677	Human angiotensino
C 26	16	13.0	50000	20	AAQ23517	Human kidney amino
C 27	15	12.2	20	22	AAQ91627	Human angiotensino
C 28	15	12.2	96	19	AAV72858	Human angiotensino
C 29	15	12.2	96	21	AAQ65347	Human angiotensino
C 30	15	12.2	101	19	AAV72819	Human angiotensino
C 31	15	12.2	101	21	AAQ65362	Human angiotensino
C 32	15	12.2	201	21	AAQ36289	Mechanical stress
C 33	15	12.2	203	21	AAQ13235	Human secreted pro
C 34	15	12.2	280	21	AAQ01492	Human colon cancer
C 35	15	12.2	300	20	AAQ14187	Human gene express
C 36	15	12.2	303	22	AAQ67192	C glutamicum codin
C 37	15	12.2	309	21	AAQ74896	Human OREX ORF451
C 38	15	12.2	319	20	AAQ40689	Human secreted pro
C 39	15	12.2	357	21	AAQ22215	Human secreted pro
C 40	15	12.2	377	22	AAQ66728	Novel human polynu
C 41	15	12.2	391	22	AAQ66301	Novel human polynu
C 42	15	12.2	412	22	AAQ66405	Novel human polynu
C 43	15	12.2	434	21	AAQ01762	Human secreted pro
C 44	15	12.2	556	21	AAQ10271	Fusarium venenatum
C 45	15	12.2	567	21	AAQ87998	Human CLAF-2 geno

#### ALIGNMENTS

RESULT 1	
AC98121	standard; cDNA; 464 BP.
XX	
AC	AAQ98121;
XX	
DT	09-MAR-2001 (first entry)
XX	
DE	Human colon cancer antigen nucleotide sequence SEQ ID NO:131.
XX	
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW	identification; cytosolic; cardiovascular; neuroprotective; vulnary;
KW	immunomodulatory; muscular; gynaecological; gastrointestinal;
KW	nephrotropic; antineoplastic; antibacterial; gene therapy; wound;
KW	neural disorder; immune system disorder; muscular disorder;
KW	reproductive disorder; gastrointestinal disorder; renal disorder;
KW	infectious disease; cardiovascular disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200055351-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000MO-US05883.
XX	
PR	12-MAR-1999; 99US-0124270.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI; 2000-587534/55.
DR	P-PDB; AAB53364.



XX Colon cancer associated gene sequences, referred to as colon cancer  
 PT antigens, useful for the treatment, prevention, and diagnosis of colon  
 PT disorders such as colon cancer -  
 XX  
 PS Claim 1; Page 565; 2104pp; English.  
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
 CC human colon cancer antigens can have cytostatic, cardiactive, muscular;  
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
 CC vulnerability, nephrotoxic, antifibrotic and antibacterial activities, and  
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
 CC proteins and antibodies to the proteins are useful for the prevention,  
 CC treatment and diagnosis of colon disorders, such as colon cancer. The  
 CC polynucleotides may be used in diagnostics and research, such as for  
 CC chromosome identification, and as hybridisation probes. The proteins  
 CC may also be used to prevent diseases such as neural disorders, immune  
 CC system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, wounds, renal disorders, infectious  
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
 CC AAB54007 represent sequences used in the exemplification of the present  
 CC invention.  
 CC  
 XX Sequence 464 BP; 112 A; 112 C; 110 G; 126 T; 4 other;  
 SO  
 Query Match 14.6%; Score 18; DB 21; Length 464;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 18 agatgccttaacaacctg 35  
 ||||||||||||||||  
 Db 324 agatgccttaacaacctg 341  
 RESULT 2  
 AAH29882/C  
 ID AAH29882 standard; DNA; 1303 BP.  
 XX  
 XX AAH29882;  
 AC  
 XX 27-JUL-2001 (first entry)  
 DT  
 XX C albicans apoptosis associated coding sequence #26.  
 DE  
 XX Yeast; fungus; apoptosis; infection; proliferative disease;  
 KW vaccine; autoimmune disease; ischaemia; neurodegeneration; ds.  
 XX  
 OS Candida albicans.  
 XX  
 PN WO200102550-A2.  
 PD  
 XX 11-JAN-2001.  
 PF 03-JUL-2000; 2000WO-BE00077.  
 XX  
 XX 01-JUL-1999; 99EP-0870141.  
 PR  
 XX (JANC ) JANSEN PHARM NV.  
 PA  
 XX Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;  
 PI Nelissen BJM, Reekmans RJ;  
 XX  
 XX WPI: 2001-367042/38.  
 DR P-PSDB; AAG70846.  
 XX  
 XX Yeast and fungal nucleic acids encoding proteins involved in a pathway  
 PT leading to programmed cell death, useful for treating proliferative  
 PT disorders, yeast and fungal infections, or for preventing apoptosis in  
 PT certain diseases -  
 XX  
 PS Claim 17; Fig 2; 218pp; English.

XX The present invention provides the protein and coding sequences of a  
 CC number of apoptosis associated proteins from the yeast *Saccharomyces*  
 CC *crevisiae* and the fungus *Candida albicans*. These can be used to identify  
 CC treatments for fungal and yeast infections, for proliferative diseases  
 CC and for apoptosis related diseases such as autoimmune diseases, ischaemia  
 CC and neurodegeneration. The present sequence is one of the *C. albicans*  
 CC coding sequences of the invention.  
 CC  
 XX Sequence 1303 BP; 488 A; 197 C; 196 G; 422 T; 0 other;  
 SO  
 Query Match 14.6%; Score 18; DB 22; Length 1303;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 106 aattattccacacaca 123  
 ||||||||||||||||  
 Db 138 AATTATTCCACACACA 121  
 RESULT 3  
 AAA81461  
 ID AAA81461 standard; DNA; 9941 BP.  
 XX  
 XX AAA81461;  
 AC  
 XX 04-DEC-2000 (first entry)  
 DT  
 XX N. meningitidis partial DNA sequence gnm\_9 SEQ ID NO:9.  
 DE  
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW *Meningococcus B*; MenB; ds.  
 XX  
 XX *Neisseria meningitidis*.  
 OS  
 XX WO200022430-A2.  
 PN  
 XX 20-APR-2000..  
 PD  
 XX 08-OCT-1999; 99WO-US23573.  
 PF  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA  
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappunli R, Pizze M;  
 XX  
 XX WPI: 2000-318079/27.  
 DR  
 XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other *Neisseria* infections, for example, *N. gonorrhoea* -  
 XX  
 PS Claim 7; Page 326-329; 1760pp; English.  
 CC  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to



CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus B*; against all serotypes;  
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 CC  
 SQ Sequence 9941 BP; 2330 A; 2662 C; 2649 G; 2300 T; 0 other;

Query Match 14.6%; Score 18; DB 21; Length 9941;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 97 ctatctccaattattc 114  
 ||||||||||||||||  
 Db 7763 ctatctccaattattc 7780

## RESULT 4

AAAF21607  
 ID AAAF21607 standard; DNA; 349980 BP.

AC AAAF21607;

DT 13-MAR-2001 (first entry)

DE *Neisseria meningitidis B* nucleotide sequence SEQ ID NO:108.

KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;  
 diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 ds.

OS *Neisseria meningitidis*.

PN WO200066791-A1.

PD 09-NOV-2000.

PF 08-MAR-2000; 2000MO-US05928.

PR 30-APR-1999; 99US-0132068.

PR 08-OCT-1999; 99WO-US23573.

PR 28-FEB-2000; 2000GB-0004695.

PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

PI Pizze M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V;

PI Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;

PI Frazer CM, Grandi G;

PI WPI; 2000-647603/62.

DR *Neisseria meningitidis B* full length genome sequence and open reading

XX frames are used to detect, treat and prevent *Neisseria* infections -

PT Claim 7; Appendix A; 692pp; English.

PS The present invention describes the full length genome of

XX *Neisseria meningitidis B* (NMB). The sequences in AAAF21544 and AAAF21607

CC to AAAF21613 represent fragments of the NMB genomic sequence, as the

CC sequence was too long to go in a record on its own it was split into 8

CC sequences which overlap each other at the beginning and end of each

CC sequence by 49980 bp (i.e. the last 49980 bp of AAAF21544 is repeated at

CC the beginning of AAAF21607, the last 49980 bp of AAAF21607 are repeated at

CC the beginning of AAAF21608, and so on). AAAF21545 to AAAF21588 encode the

CC *Neisseria* proteins given in AAB58550 to AAB58593, and AAAF21589 to  
 CC AAAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to *Neisseria* bacteria or as a diagnostic reagent for detecting the  
 CC presence of *Neisseria* bacteria or of antibodies raised to *Neisseria*  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.  
 CC  
 SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Query Match 14.6%; Score 18; DB 21; Length 349980;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 97 ctatctccaattattc 114  
 ||||||||||||||||  
 Db 103322 ctatctccaattattc 103339

## RESULT 5

AAA81490  
 ID AAA81490 standard; DNA; 1437668 BP.

AC AAA81490;

DT 04-DEC-2000 (first entry)

DE *N. meningitidis B* full length genome DNA sequence SEQ ID NO:1068.

KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;  
 antigen; vaccine; diagnosis; infection; antibacterial; identification;  
*Meningococcus B*; MenB; ds.

OS *Neisseria meningitidis*.

PN WO200022430-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99WO-US23573.

PR 09-OCT-1998; 98US-0103794.

PR 30-APR-1999; 99US-0132068.

PA (CHIR) CHIRON CORP.

PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;

PI Rappuoli R, Pizze M;

PI WPI; 2000-318079/27.

DR Isolated nucleotide sequences of *Neisseria meningitidis* which can be

XX used in the diagnosis and treatment of *N. meningitidis* infection and

PT other *Neisseria* infections, for example, *N. gonorrhoea* -

PS Claim 7; Page 866-1272; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic

CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414

CC represent specifically claimed *Neisseria meningitidis* genomic DNA

CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent

CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to

CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the

CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to



CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medication (or in the manufacture of a  
 CC medication) for treating, preventing or diagnosing infection due to  
 CC *Neisseria* bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus B*; against all serotypes;  
 CC and/or against all pathogenic *Neisseria*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other:  
 SQ

Query Match 14.6%; Score 18; DB 21; Length 1437668;  
 Best Local Similarity 100.0%; Pred. No. 1.9; Mismatches 0; Indels 0; Gaps 0;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 97 ctatcttcattatc 114  
 |||||  
 Db 403322 ctatcttcattatc 403339

RESULT 6  
 AAH29434/C  
 ID AAH29434 standard; DNA; 518 BP.  
 XX  
 AC AAH29434;  
 XX  
 DT 17-JUL-2001 (first entry)  
 XX  
 DE *Drosophila melanogaster* essential gene fragment, SEQ ID NO: 623.  
 XX  
 KM *Drosophila melanogaster*; fruit fly; essential gene; screening assay;  
 KW pesticide; crop protection; chromosome 3; ds.  
 XX  
 OS *Drosophila melanogaster*.  
 XX  
 PN WO200118547-A1.  
 XX  
 PD 15-MAR-2001.  
 XX  
 PF 06-SEP-2000; 2000WO-GB03444.  
 XX  
 PR 07-SEP-1999; 99GB-0021009.  
 XX  
 PA (UNIT) UNIT GLASGOW.  
 XX  
 PI Davies RW, Kaiser K, Yang MY;  
 XX  
 DR WPI; 2001-281436/29.  
 XX  
 PT Screening assays for used for identifying compounds having a  
 PT physiological effect on proteins identified as being essential -  
 XX  
 PS Claim 23; Page 519; 695pp; English.  
 XX  
 CC The present sequence is part of an essential gene from *Drosophila*  
 CC *melanogaster*. Lack of expression of the protein encoded by this  
 CC gene leads to a lethal or semi-lethal phenotype. The invention  
 CC relates to 902 nucleic acid sequences from genes encoding proteins  
 CC which are thought to be essential, and to a screening assay for  
 CC identifying compounds which have a physiological effect on these  
 CC proteins. Suitable compounds are useful as pesticides and may be used  
 CC in conjunction with other pesticides and herbicides for crop

CC protection. The gene corresponding to the present sequence is located  
 CC on chromosome 3.  
 CC  
 SQ Sequence 518 BP; 134 A; 127 C; 117 G; 140 T; 0 other;

Query Match 13.8%; Score 17; DB 22; Length 518;  
 Best Local Similarity 100.0%; Pred. No. 7.6;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 50 caatgcaggcgtgcac 66  
 |||||  
 Db 208 CAATGCAGGCGTGCAC 192

RESULT 7  
 AA103888/C  
 ID AA103888 standard; DNA; 446 BP.  
 XX  
 AC AA103888;  
 XX  
 DT 09-OCT-2001 (first entry)  
 XX  
 DE Probe #3879 used to measure gene expression in human breast sample.  
 XX  
 KW Probe; human; breast disease; breast cancer; development disorder; ss;  
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 OS *Homo sapiens*.  
 XX  
 PN WO200157270-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 29-JAN-2001; 2001WO-US00661.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-476286/51.  
 XX  
 PT Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 XX  
 PS Claim 25; SEQ ID NO 3879; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes.  
 CC The present sequence is one such probe. The probes are useful for  
 CC measuring human gene expression in a human breast sample, where the probe  
 CC hybridises at high stringency to a nucleic acid expressed in the human  
 CC breast. The probes are useful for predicting, diagnosing, grading,  
 CC staging, monitoring and prognosing diseases of the human breast,  
 CC particularly those diseases with polygenic aetiology. The diseases  
 CC include: breast cancer, disorders of development, inflammatory diseases  
 CC of the breast, fibrocystic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 446 BP; 126 A; 98 C; 83 G; 139 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 446;



Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 taagtccacatgtaa 89  
| | | | | | | | | | | | | | | |  
Db 160 TAAGTTTCACATGTAA 145

## RESULT 8

AAI14349/C  
ID AAI14349 standard; DNA: 459 BP.

XX  
AC AAI14349;

XX  
DT 12-OCT-2001 (first entry)

DE Probe #4282 for gene expression analysis in human cervical cell sample.

XX  
KW Probe: human; microarray; gene expression; cervical epithelial cell;

XX  
KM cervical cancer; ss.

XX  
OS Homo sapiens.

XX  
PN WO200157278-A2.

XX  
PD . 09-AUG-2001.

XX  
PF 30-JAN-2001; 2001WO-US00670.

XX  
PR 04-FEB-2000; 2000US-0180312.

XX  
PR 26-MAY-2000; 2000US-0207456.

XX  
PR 30-JUN-2000; 2000US-0608408.

XX  
PR 03-AUG-2000; 2000US-0632366.

XX  
PR 21-SEP-2000; 2000US-0234687.

XX  
PR 27-SEP-2000; 2000US-0236359.

XX  
PR 04-OCT-2000; 2000GB-0024263.

XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX  
DR WPI; 2001-488901/53.

XX  
PT Human genome-derived single exon nucleic acid probes useful for

XX  
PS analyzing gene expression in human cervical epithelial cells -

XX  
PS Claim 25; SEQ ID No 4282; 487pp; English.

XX  
CC The present invention relates to human single exon nucleic acid probes

XX  
CC (SENP). The present sequence is one such probe. The SENPs are derived

XX  
CC from human Hela cells. The SENPs can be used to produce a single exon

XX  
CC microarray, which can be used for measuring human gene expression in a

XX  
CC sample derived from human cervical epithelial cells. By measuring gene

XX  
CC expression, the probes are therefore useful in grading and/or staging

XX  
CC of diseases of the cervix, notably cervical cancer.

XX  
CC Note: The sequence data for this patent did not form part of the printed

XX  
CC specification, but was obtained in electronic format directly from WIPO

XX  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 13.0%; Score 16; DB 22; Length 459;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ccctgatgacctaatg 46  
| | | | | | | | | | | | | | | |  
Db 409 CCCTGATGACCTAATG 394

## RESULT 9

AAI35723/C

ID AAI35723 standard; DNA: 459 BP.

XX  
AC AAI35723;

XX  
DT 17-OCT-2001 (first entry)

DE Probe #4409 used to measure gene expression in human placenta sample.

XX  
KW Probe: microarray; human; placenta; antenatal diagnosis;

XX  
KM genetic disorder; ss.

XX  
OS Homo sapiens.

XX  
PN WO200157272-A2.

XX  
PD 09-AUG-2001.

XX  
PF 30-JAN-2001; 2001WO-US00663.

XX  
PR 04-FEB-2000; 2000US-0180312.

XX  
PR 26-MAY-2000; 2000US-0207456.

XX  
PR 30-JUN-2000; 2000US-0608408.

XX  
PR 03-AUG-2000; 2000US-0632366.

XX  
PR 21-SEP-2000; 2000US-0234687.

XX  
PR 27-SEP-2000; 2000US-0236359.

XX  
PR 04-OCT-2000; 2000GB-0024263.

XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX  
DR WPI; 2001-48897/53.

XX  
PT Human genome-derived single exon nucleic acid probes useful for

XX  
PS analyzing gene expression in human placenta -

XX  
PS Claim 25; SEQ ID No 4409; 654pp; English.

XX  
CC The present invention relates to single exon nucleic acid probes (SENP).

XX  
CC The present sequence is one such probe. The probes are useful for

XX  
CC producing a microarray for predicting, measuring and displaying gene

XX  
CC expression in samples derived from human placenta. The probes are useful

XX  
CC for antenatal diagnosis of human genetic disorders.

XX  
CC Sequence 459 BP; 137 A; 61 C; 110 G; 151 T; 0 other;

XX  
CC Query Match 13.0%; Score 16; DB 22; Length 459;

XX  
CC Best Local Similarity 100.0%; Pred. No. 26;

XX  
CC Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ccctgatgacctaatg 46  
| | | | | | | | | | | | | | | |  
Db 409 CCCTGATGACCTAATG 394

## RESULT 10

AAI04179/C

ID AAI04179 standard; DNA: 459 BP.

XX  
AC AAI04179;

XX  
DT 09-OCT-2001 (first entry)

DE Probe #4170 used to measure gene expression in human breast sample.

XX  
KW Probe: human; breast disease; breast cancer; development disorder; ss;

XX  
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX  
OS Homo sapiens.

XX  
PN WO200157270-A2.



PD 09-AUG-2001.  
 XX  
 XX 29-JAN-2001; 2001WO-US000661.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-476286/51.  
 XX  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 PS  
 PS Claim 25; SEQ ID No 4170; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes.  
 CC The present sequence is one such probe. The probes are useful for  
 CC measuring human gene expression in a human breast sample, where the probe  
 CC hybridizes at high stringency to a nucleic acid expressed in the human  
 CC breast. The probes are useful for predicting, diagnosing, grading,  
 CC staging, monitoring and prognosing diseases of the human breast,  
 CC particularly those diseases with polygenic aetiology. The diseases  
 CC include: breast cancer, disorders of development, inflammatory diseases  
 CC of the breast, fibrocystic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 459 BP; 137 A; 61 C; 110 G; 151 T; 0 other:  
 SQ

Query Match 13.0%; Score 16; DB 22; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 31 cccgtatgacctaatg 46  
 ||||||||||||  
 DB 409 CCTGATGACTAATG 394

RESULT 11  
 AAX34755/c  
 ID AAX34755 standard; DNA; 852 BP.  
 XX  
 AC AAX34755;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE DNA encoding OMP-1W protein.  
 XX  
 KW Outer membrane protein; OMP; *Escherichia chaffeensis*; *E. canis*; P30;  
 KW detection; dog; ss.  
 XX  
 OS *Escherichia chaffeensis*.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.

XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 DR P-Psdb; AAY06955.  
 XX  
 XX Novel outer membrane proteins from *Escherichia chaffeensis* and  
 PT *Escherichia canis*  
 PS  
 PS Disclosure; Fig 15A; 55pp; English.  
 CC  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC *Escherichia chaffeensis* and *E. canis*. The *E. chaffeensis* proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The *E. canis* proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect *E. chaffeensis* in patients and *E. canis* in dogs.  
 XX  
 XX Sequence 852 BP; 295 A; 115 C; 141 G; 301 T; 0 other;  
 SQ

Query Match 13.0%; Score 16; DB 20; Length 852;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 84 atgaagtatactt 99  
 ||||||||||||  
 DB 578 ATGTAAGTGCTACTT 563

RESULT 12  
 AAX38323/c  
 ID AAX38323 standard; DNA; 1278 BP.  
 XX  
 AC AAX38323;  
 XX  
 DT 21-AUG-2000 (first entry)  
 XX  
 DE Human angiotensinogen (AGT) gene regulatory sequence (GenBank X15323).  
 DE  
 KW Angiotensinogen gene; AGT; regulatory region; polymorphism;  
 KW polymorphic marker; cardiovascular disease; myocardial infarction;  
 KW unstable angina; hypertension; atherosclerosis; stroke; prognosis;  
 KW drug screening; treatment outcome; human; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200022166-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 13-OCT-1999; 99WO-IB01678.  
 XX  
 PR 14-OCT-1998; 98US-0104286.  
 PR 14-OCT-1998; 98US-0104302.  
 XX  
 PA (EURO-) EUROPA MEDICAL AB.  
 XX  
 PI Norberg LT, Andersson MK, Lindstrom PHR, Jonsson L;  
 XX  
 DR WPI; 2000-318010/27.  
 XX  
 PT Assessing cardiovascular status in humans involves comparing test  
 PT polymorphic pattern comprising polymorphic positions within genes  
 PT encoding specific proteins, with reference polymorphic pattern -  
 XX  
 PS Disclosure; Page -; 126pp; English.  
 XX  
 CC The invention relates to a novel method of assessing the cardiovascular  
 CC status in an individual and to newly identified polymorphisms in the  
 CC genes encoding angiotensin-converting enzyme (ACE), angiotensin II  
 CC receptor type 1 (AT1) and type 2 (AT2), angiotensinogen (AGT), renin,  
 CC aldosterone synthase, endothelin receptor type A and beta-adrenergic



receptors 1 and 2. The method comprises determining the sequence at one or more polymorphic positions within these genes, and comparing the pattern of polymorphisms from the individual with a reference polymorphic pattern obtained from a population of individuals exhibiting a predetermined cardiovascular disease status. The polymorphic markers are useful for determining the predisposition of an individual to cardiovascular disorders such as myocardial infarction, unstable angina, hypertension, atherosclerosis and stroke. They are also useful for predicting the likely cardiovascular status of a patient given a treatment regimen comprising administration of cardiovascular drugs (e.g., ACE inhibitors, beta-adrenergic receptor antagonists (beta-blockers) or calcium channel blockers). One or more polymorphic markers provides a basis for predicting the outcome of a treatment regimen. Fragments of the genes comprising a polymorphic site may be used as primers and probes for detecting genetic polymorphisms or in molecular library arrays for high throughput screening. The genes, and the proteins they encode are useful in the screening of potential cardiovascular drugs. Determination of an individual's polymorphic pattern reduces or eliminates trial and error in selecting a treatment for a particular individual cardiovascular patient. It also provides the ability to eliminate patients from clinical trials who are predicted to be non-responsive, or at a risk for an adverse response, to a particular treatment regimen. Adverse results in an early trial can be evaluated to identify polymorphic patterns so that the adverse results can be correlated with a sub-population of the test population, permitting exclusion of such sub-populations from the treatment group. Beneficial drugs can be approved for use in the appropriate population, thereby decreasing the number of patients required for a clinical trial, which in turn decreases the duration and cost of such trials. The present sequence represents the human angiotensinogen gene regulatory sequence (GenBank X15333). The polymorphic sites identified are 3957T/A, 412C/T, 432G/A, 449T/C, 692C/T, 839G/A, 1007G/A, 1072G/A, 1204C/A, 1218A/G, 1240C/A, 1241G/A, 1242G/A, 1243G/A, 1244G/A, 1245G/A, 1246G/A, 1247G/A, 1248G/A, 1249G/A, 1250G/A, 1251G/A, 1252G/A, 1253G/A, 1254G/A, 1255G/A, 1256G/A, 1257G/A, 1258G/A, 1259G/A, 1260G/A, 1261G/A, 1262G/A, 1263G/A, 1264G/A, 1265G/A, 1266G/A, 1267G/A, 1268G/A, 1269G/A, 1270G/A, 1271G/A, 1272G/A, 1273G/A, 1274G/A, 1275G/A, 1276G/A, 1277G/A, 1278G/A, 1279G/A, 1280G/A, 1281G/A, 1282G/A, 1283G/A, 1284G/A, 1285G/A, 1286G/A, 1287G/A, 1288G/A, 1289G/A, 1290G/A, 1291G/A, 1292G/A, 1293G/A, 1294G/A, 1295G/A, 1296G/A, 1297G/A, 1298G/A, 1299G/A, 1300G/A, 1301G/A, 1302G/A, 1303G/A, 1304G/A, 1305G/A, 1306G/A, 1307G/A, 1308G/A, 1309G/A, 1310G/A, 1311G/A, 1312G/A, 1313G/A, 1314G/A, 1315G/A, 1316G/A, 1317G/A, 1318G/A, 1319G/A, 1320G/A, 1321G/A, 1322G/A, 1323G/A, 1324G/A, 1325G/A, 1326G/A, 1327G/A, 1328G/A, 1329G/A, 1330G/A, 1331G/A, 1332G/A, 1333G/A, 1334G/A, 1335G/A, 1336G/A, 1337G/A, 1338G/A, 1339G/A, 1340G/A, 1341G/A, 1342G/A, 1343G/A, 1344G/A, 1345G/A, 1346G/A, 1347G/A, 1348G/A, 1349G/A, 1350G/A, 1351G/A, 1352G/A, 1353G/A, 1354G/A, 1355G/A, 1356G/A, 1357G/A, 1358G/A, 1359G/A, 1360G/A, 1361G/A, 1362G/A, 1363G/A, 1364G/A, 1365G/A, 1366G/A, 1367G/A, 1368G/A, 1369G/A, 1370G/A, 1371G/A, 1372G/A, 1373G/A, 1374G/A, 1375G/A, 1376G/A, 1377G/A, 1378G/A, 1379G/A, 1380G/A, 1381G/A, 1382G/A, 1383G/A, 1384G/A, 1385G/A, 1386G/A, 1387G/A, 1388G/A, 1389G/A, 1390G/A, 1391G/A, 1392G/A, 1393G/A, 1394G/A, 1395G/A, 1396G/A, 1397G/A, 1398G/A, 1399G/A, 1400G/A, 1401G/A, 1402G/A, 1403G/A, 1404G/A, 1405G/A, 1406G/A, 1407G/A, 1408G/A, 1409G/A, 1410G/A, 1411G/A, 1412G/A, 1413G/A, 1414G/A, 1415G/A, 1416G/A, 1417G/A, 1418G/A, 1419G/A, 1420G/A, 1421G/A, 1422G/A, 1423G/A, 1424G/A, 1425G/A, 1426G/A, 1427G/A, 1428G/A, 1429G/A, 1430G/A, 1431G/A, 1432G/A, 1433G/A, 1434G/A, 1435G/A, 1436G/A, 1437G/A, 1438G/A, 1439G/A, 1440G/A, 1441G/A, 1442G/A, 1443G/A, 1444G/A, 1445G/A, 1446G/A, 1447G/A, 1448G/A, 1449G/A, 1450G/A, 1451G/A, 1452G/A, 1453G/A, 1454G/A, 1455G/A, 1456G/A, 1457G/A, 1458G/A, 1459G/A, 1460G/A, 1461G/A, 1462G/A, 1463G/A, 1464G/A, 1465G/A, 1466G/A, 1467G/A, 1468G/A, 1469G/A, 1470G/A, 1471G/A, 1472G/A, 1473G/A, 1474G/A, 1475G/A, 1476G/A, 1477G/A, 1478G/A, 1479G/A, 1480G/A, 1481G/A, 1482G/A, 1483G/A, 1484G/A, 1485G/A, 1486G/A, 1487G/A, 1488G/A, 1489G/A, 1490G/A, 1491G/A, 1492G/A, 1493G/A, 1494G/A, 1495G/A, 1496G/A, 1497G/A, 1498G/A, 1499G/A, 1500G/A, 1501G/A, 1502G/A, 1503G/A, 1504G/A, 1505G/A, 1506G/A, 1507G/A, 1508G/A, 1509G/A, 1510G/A, 1511G/A, 1512G/A, 1513G/A, 1514G/A, 1515G/A, 1516G/A, 1517G/A, 1518G/A, 1519G/A, 1520G/A, 1521G/A, 1522G/A, 1523G/A, 1524G/A, 1525G/A, 1526G/A, 1527G/A, 1528G/A, 1529G/A, 1530G/A, 1531G/A, 1532G/A, 1533G/A, 1534G/A, 1535G/A, 1536G/A, 1537G/A, 1538G/A, 1539G/A, 1540G/A, 1541G/A, 1542G/A, 1543G/A, 1544G/A, 1545G/A, 1546G/A, 1547G/A, 1548G/A, 1549G/A, 1550G/A, 1551G/A, 1552G/A, 1553G/A, 1554G/A, 1555G/A, 1556G/A, 1557G/A, 1558G/A, 1559G/A, 1560G/A, 1561G/A, 1562G/A, 1563G/A, 1564G/A, 1565G/A, 1566G/A, 1567G/A, 1568G/A, 1569G/A, 1570G/A, 1571G/A, 1572G/A, 1573G/A, 1574G/A, 1575G/A, 1576G/A, 1577G/A, 1578G/A, 1579G/A, 1580G/A, 1581G/A, 1582G/A, 1583G/A, 1584G/A, 1585G/A, 1586G/A, 1587G/A, 1588G/A, 1589G/A, 1590G/A, 1591G/A, 1592G/A, 1593G/A, 1594G/A, 1595G/A, 1596G/A, 1597G/A, 1598G/A, 1599G/A, 1600G/A, 1601G/A, 1602G/A, 1603G/A, 1604G/A, 1605G/A, 1606G/A, 1607G/A, 1608G/A, 1609G/A, 1610G/A, 1611G/A, 1612G/A, 1613G/A, 1614G/A, 1615G/A, 1616G/A, 1617G/A, 1618G/A, 1619G/A, 1620G/A, 1621G/A, 1622G/A, 1623G/A, 1624G/A, 1625G/A, 1626G/A, 1627G/A, 1628G/A, 1629G/A, 1630G/A, 1631G/A, 1632G/A, 1633G/A, 1634G/A, 1635G/A, 1636G/A, 1637G/A, 1638G/A, 1639G/A, 1640G/A, 1641G/A, 1642G/A, 1643G/A, 1644G/A, 1645G/A, 1646G/A, 1647G/A, 1648G/A, 1649G/A, 1650G/A, 1651G/A, 1652G/A, 1653G/A, 1654G/A, 1655G/A, 1656G/A, 1657G/A, 1658G/A, 1659G/A, 1660G/A, 1661G/A, 1662G/A, 1663G/A, 1664G/A, 1665G/A, 1666G/A, 1667G/A, 1668G/A, 1669G/A, 1670G/A, 1671G/A, 1672G/A, 1673G/A, 1674G/A, 1675G/A, 1676G/A, 1677G/A, 1678G/A, 1679G/A, 1680G/A, 1681G/A, 1682G/A, 1683G/A, 1684G/A, 1685G/A, 1686G/A, 1687G/A, 1688G/A, 1689G/A, 1690G/A, 1691G/A, 1692G/A, 1693G/A, 1694G/A, 1695G/A, 1696G/A, 1697G/A, 1698G/A, 1699G/A, 1700G/A, 1701G/A, 1702G/A, 1703G/A, 1704G/A, 1705G/A, 1706G/A, 1707G/A, 1708G/A, 1709G/A, 1710G/A, 1711G/A, 1712G/A, 1713G/A, 1714G/A, 1715G/A, 1716G/A, 1717G/A, 1718G/A, 1719G/A, 1720G/A, 1721G/A, 1722G/A, 1723G/A, 1724G/A, 1725G/A, 1726G/A, 1727G/A, 1728G/A, 1729G/A, 1730G/A, 1731G/A, 1732G/A, 1733G/A, 1734G/A, 1735G/A, 1736G/A, 1737G/A, 1738G/A, 1739G/A, 1740G/A, 1741G/A, 1742G/A, 1743G/A, 1744G/A, 1745G/A, 1746G/A, 1747G/A, 1748G/A, 1749G/A, 1750G/A, 1751G/A, 1752G/A, 1753G/A, 1754G/A, 1755G/A, 1756G/A, 1757G/A, 1758G/A, 1759G/A, 1760G/A, 1761G/A, 1762G/A, 1763G/A, 1764G/A, 1765G/A, 1766G/A, 1767G/A, 1768G/A, 1769G/A, 1770G/A, 1771G/A, 1772G/A, 1773G/A, 1774G/A, 1775G/A, 1776G/A, 1777G/A, 1778G/A, 1779G/A, 1780G/A, 1781G/A, 1782G/A, 1783G/A, 1784G/A, 1785G/A, 1786G/A, 1787G/A, 1788G/A, 1789G/A, 1790G/A, 1791G/A, 1792G/A, 1793G/A, 1794G/A, 1795G/A, 1796G/A, 1797G/A, 1798G/A, 1799G/A, 1800G/A, 1801G/A, 1802G/A, 1803G/A, 1804G/A, 1805G/A, 1806G/A, 1807G/A, 1808G/A, 1809G/A, 1810G/A, 1811G/A, 1812G/A, 1813G/A, 1814G/A, 1815G/A, 1816G/A, 1817G/A, 1818G/A, 1819G/A, 1820G/A, 1821G/A, 1822G/A, 1823G/A, 1824G/A, 1825G/A, 1826G/A, 1827G/A, 1828G/A, 1829G/A, 1830G/A, 1831G/A, 1832G/A, 1833G/A, 1834G/A, 1835G/A, 1836G/A, 1837G/A, 1838G/A, 1839G/A, 1840G/A, 1841G/A, 1842G/A, 1843G/A, 1844G/A, 1845G/A, 1846G/A, 1847G/A, 1848G/A, 1849G/A, 1850G/A, 1851G/A, 1852G/A, 1853G/A, 1854G/A, 1855G/A, 1856G/A, 1857G/A, 1858G/A, 1859G/A, 1860G/A, 1861G/A, 1862G/A, 1863G/A, 1864G/A, 1865G/A, 1866G/A, 1867G/A, 1868G/A, 1869G/A, 1870G/A, 1871G/A, 1872G/A, 1873G/A, 1874G/A, 1875G/A, 1876G/A, 1877G/A, 1878G/A, 1879G/A, 1880G/A, 1881G/A, 1882G/A, 1883G/A, 1884G/A, 1885G/A, 1886G/A, 1887G/A, 1888G/A, 1889G/A, 1890G/A, 1891G/A, 1892G/A, 1893G/A, 1894G/A, 1895G/A, 1896G/A, 1897G/A, 1898G/A, 1899G/A, 1900G/A, 1901G/A, 1902G/A, 1903G/A, 1904G/A, 1905G/A, 1906G/A, 1907G/A, 1908G/A, 1909G/A, 1910G/A, 1911G/A, 1912G/A, 1913G/A, 1914G/A, 1915G/A, 1916G/A, 1917G/A, 1918G/A, 1919G/A, 1920G/A, 1921G/A, 1922G/A, 1923G/A, 1924G/A, 1925G/A, 1926G/A, 1927G/A, 1928G/A, 1929G/A, 1930G/A, 1931G/A, 1932G/A, 1933G/A, 1934G/A, 1935G/A, 1936G/A, 1937G/A, 1938G/A, 1939G/A, 1940G/A, 1941G/A, 1942G/A, 1943G/A, 1944G/A, 1945G/A, 1946G/A, 1947G/A, 1948G/A, 1949G/A, 1950G/A, 1951G/A, 1952G/A, 1953G/A, 1954G/A, 1955G/A, 1956G/A, 1957G/A, 1958G/A, 1959G/A, 1960G/A, 1961G/A, 1962G/A, 1963G/A, 1964G/A, 1965G/A, 1966G/A, 1967G/A, 1968G/A, 1969G/A, 1970G/A, 1971G/A, 1972G/A, 1973G/A, 1974G/A, 1975G/A, 1976G/A, 1977G/A, 1978G/A, 1979G/A, 1980G/A, 1981G/A, 1982G/A, 1983G/A, 1984G/A, 1985G/A, 1986G/A, 1987G/A, 1988G/A, 1989G/A, 1990G/A, 1991G/A, 1992G/A, 1993G/A, 1994G/A, 1995G/A, 1996G/A, 1997G/A, 1998G/A, 1999G/A, 2000G/A, 2001G/A, 2002G/A, 2003G/A, 2004G/A, 2005G/A, 2006G/A, 2007G/A, 2008G/A, 2009G/A, 2010G/A, 2011G/A, 2012G/A, 2013G/A, 2014G/A, 2015G/A, 2016G/A, 2017G/A, 2018G/A, 2019G/A, 2020G/A, 2021G/A, 2022G/A, 2023G/A, 2024G/A, 2025G/A, 2026G/A, 2027G/A, 2028G/A, 2029G/A, 2030G/A, 2031G/A, 2032G/A, 2033G/A, 2034G/A, 2035G/A, 2036G/A, 2037G/A, 2038G/A, 2039G/A, 2040G/A, 2041G/A, 2042G/A, 2043G/A, 2044G/A, 2045G/A, 2046G/A, 2047G/A, 2048G/A, 2049G/A, 2050G/A, 2051G/A, 2052G/A, 2053G/A, 2054G/A, 2055G/A, 2056G/A, 2057G/A, 2058G/A, 2059G/A, 2060G/A, 2061G/A, 2062G/A, 2063G/A, 2064G/A, 2065G/A, 2066G/A, 2067G/A, 2068G/A, 2069G/A, 2070G/A, 2071G/A, 2072G/A, 2073G/A, 2074G/A, 2075G/A, 2076G/A, 2077G/A, 2078G/A, 2079G/A, 2080G/A, 2081G/A, 2082G/A, 2083G/A, 2084G/A, 2085G/A, 2086G/A, 2087G/A, 2088G/A, 2089G/A, 2090G/A, 2091G/A, 2092G/A, 2093G/A, 2094G/A, 2095G/A, 2096G/A, 2097G/A, 2098G/A, 2099G/A, 2100G/A, 2101G/A, 2102G/A, 2103G/A, 2104G/A, 2105G/A, 2106G/A, 2107G/A, 2108G/A, 2109G/A, 2110G/A, 2111G/A, 2112G/A, 2113G/A, 2114G/A, 2115G/A, 2116G/A, 2117G/A, 2118G/A, 2119G/A, 2120G/A, 2121G/A, 2122G/A, 2123G/A, 2124G/A, 2125G/A, 2126G/A, 2127G/A, 2128G/A, 2129G/A, 2130G/A, 2131G/A, 2132G/A, 2133G/A, 2134G/A, 2135G/A, 2136G/A, 2137G/A, 2138G/A, 2139G/A, 2140G/A, 2141G/A, 2142G/A, 2143G/A, 2144G/A, 2145G/A, 2146G/A, 2147G/A, 2148G/A, 2149G/A, 2150G/A, 2151G/A, 2152G/A, 2153G/A, 2154G/A, 2155G/A, 2156G/A, 2157G/A, 2158G/A, 2159G/A, 2160G/A, 2161G/A, 2162G/A, 2163G/A, 2164G/A, 2165G/A, 2166G/A, 2167G/A, 2168G/A, 2169G/A, 2170G/A, 2171G/A, 2172G/A, 2173G/A, 2174G/A, 2175G/A, 2176G/A, 2177G/A, 2178G/A, 2179G/A, 2180G/A, 2181G/A, 2182G/A, 2183G/A, 2184G/A, 2185G/A, 2186G/A, 2187G/A, 2188G/A, 2189G/A, 2190G/A, 2191G/A, 2192G/A, 2193G/A, 2194G/A, 2195G/A, 2196G/A, 2197G/A, 2198G/A, 2199G/A, 2200G/A, 2201G/A, 2202G/A, 2203G/A, 2204G/A, 2205G/A, 2206G/A, 2207G/A, 2208G/A, 2209G/A, 2210G/A, 2211G/A, 2212G/A, 2213G/A, 2214G/A, 2215G/A, 2216G/A, 2217G/A, 2218G/A, 2219G/A, 2220G/A, 2221G/A, 2222G/A, 2223G/A, 2224G/A, 2225G/A, 2226G/A, 2227G/A, 2228G/A, 2229G/A, 2230G/A, 2231G/A, 2232G/A, 2233G/A, 2234G/A, 2235G/A, 2236G/A, 2237G/A, 2238G/A, 2239G/A, 2240G/A, 2241G/A, 2242G/A, 2243G/A, 2244G/A, 2245G/A, 2246G/A, 2247G/A, 2248G/A, 2249G/A, 2250G/A, 2251G/A, 2252G/A, 2253G/A, 2254G/A, 2255G/A, 2256G/A, 2257G/A, 2258G/A, 2259G/A, 2260G/A, 2261G/A, 2262G/A, 2263G/A, 2264G/A, 2265G/A, 2266G/A, 2267G/A, 2268G/A, 2269G/A, 2270G/A, 2271G/A, 2272G/A, 2273G/A, 2274G/A, 2275G/A, 2276G/A, 2277G/A, 2278G/A, 2279G/A, 2280G/A, 2281G/A, 2282G/A, 2283G/A, 2284G/A, 2285G/A, 2286G/A, 2287G/A, 2288G/A, 2289G/A, 2290G/A, 2291G/A, 2292G/A, 2293G/A, 2294G/A, 2295G/A, 2296G/A, 2297G/A, 2298G/A, 2299G/A, 2300G/A, 2301G/A, 2302G/A, 2303G/A, 2304G/A, 2305G/A, 2306G/A, 2307G/A, 2308G/A, 2309G/A, 2310G/A, 2311G/A, 2312G/A, 2313G/A, 2314G/A, 2315G/A, 2316G/A, 2317G/A, 2318G/A, 2319G/A, 2320G/A, 2321G/A, 2322G/A, 2323G/A, 2324G/A, 2325G/A, 2326G/A, 2327G/A, 2328G/A, 2329G/A, 2330G/A, 2331G/A, 2332G/A, 2333G/A, 2334G/A, 2335G/A, 2336G/A, 2337G/A, 2338G/A, 2339G/A, 2340G/A, 2341G/A, 2342G/A, 2343G/A, 2344G/A, 2345G/A, 2346G/A, 2347G/A, 2348G/A, 2349G/A, 2350G/A, 2351G/A, 2352G/A, 2353G/A, 2354G/A, 2355G/A, 2356G/A, 2357G/A, 2358G/A, 2359G/A, 2360G/A, 2361G/A, 2362G/A, 2363G/A, 2364G/A, 2365G/A, 2366G/A, 2367G/A, 2368G/A, 2369G/A, 2370G/A, 2371G/A, 2372G/A, 2373G/A, 2374G/A, 2375G/A, 2376G/A, 2377G/A, 2378G/A, 2379G/A, 2380G/A, 2381G/A, 2382G/A, 2383G/A, 2384G/A, 2385G/A, 2386G/A, 2387G/A, 2388G/A, 2389G/A, 2390G/A, 2391G/A, 2392G/A, 2393G/A, 2394G/A, 2395G/A, 2396G/A, 2397G/A, 2398G/A, 2399G/A, 2400G/A, 2401G/A, 2402G/A, 2403G/A, 2404G/A, 2405G/A, 2406G/A, 2407G/A, 2408G/A, 2409G/A, 2410G/A, 2411G/A, 2412G/A, 2413G/A, 2414G/A, 2415G/A, 2416G/A, 2417G/A, 2418G/A, 2419G/A, 2420G/A, 2421G/A, 2422G/A, 2423G/A, 2424G/A, 2425G/A, 2426G/A, 2427G/A, 2428G/A, 2429G/A, 2430G/A, 2431G/A, 2432G/A, 2433G/A, 2434G/A, 2435G/A, 2436G/A, 2437G/A, 2438G/A, 2439G/A, 2440G/A, 2441G/A, 2442G/A, 2443G/A, 2444G/A, 2445G/A, 2446G/A, 2447G/A, 2448G/A, 2449G/A, 2450G/A, 2451G/A, 2452G/A, 2453G/A, 2454G/A, 2455G/A, 2456G/A, 2457G/A, 2458G/A, 2459G/A, 2460G/A, 2461G/A, 2462G/A, 2463G/A, 2464G/A, 2465G/A, 2466G/A, 2467G/A, 2468G/A, 2469G/A, 2470G/A, 2471G/A, 2472G/A, 2473G/A, 2474G/A, 2475G/A, 2476G/A, 2477G/A, 2478G/A, 2479G/A, 2480G/A, 2481G/A, 2482G/A, 2483G/A, 2484G/A, 2485G/A, 2486G/A, 2487G/A, 2488G/A, 2489G/A, 2490G/A, 2491G/A, 2492G/A, 2493G/A, 2494G/A, 2495G/A, 2496G/A, 2497G/A, 2498G/A, 2499G/A, 2500G/A, 2501G/A, 2502G/A, 2503G/A, 2504G/A, 2505G/A, 2506G/A, 2507G/A, 2508G/A, 2509G/A, 2510G/A, 2511G/A, 2512G/A, 2513G/A, 2514G/A, 2515G/A, 2516G/A, 2517G/A, 2518G/A, 2519G/A, 2520G/A, 2521G/A, 2522G/A, 2523G/A, 2524G/A, 2525G/A, 2526G/A, 2527G/A, 2528G/A, 2529G/A, 2530G/A, 2531G/A, 2532G/A, 2533G/A, 2534G/A, 2535G/A, 2536G/A, 2537G/A, 2538G/A, 2539G/A, 2540G/A, 2541G/A, 2542G/A, 2543G/A, 2544G/A, 2545G/A, 2546G/A, 2547G/A, 2548G/A, 2549G/A, 2550G/A, 2551G/A, 2552G/A, 2553G/A, 2554G/A, 2555G/A, 2556G/A, 2557G/A, 2558G/A, 2559G/A, 2560G/A, 2561G/A, 2562G/A, 2563G/A, 2564G/A, 2565G/A, 2566G/A, 2567G/A, 2568G/A, 2569G/A, 2570G/A, 2571G/A, 2572G/A, 2573G/A, 2574G/A, 2575G/A, 2576G/A, 2577G/A, 2578G/A, 2579G/A, 2580G/A, 2581G/A, 2582G/A, 2583G/A, 2584G/A, 2585G/A, 2586G/A, 2587G/A, 2588G/A, 2589G/A, 2590G/A, 2591G/A, 2592G/A, 2593G/A, 2594G/A, 2595G/A, 2596G/A, 2597G/A, 2598G/A, 2599G/A, 2600G/A, 2601G/A, 2602G/A, 2603G/A, 2604G/A, 2605G/A, 2606G/A, 2607G/A, 2608G/A, 2609G/A, 2610G/A, 2611G/A, 2612G/A, 2613G/A, 2614G/A, 2615G/A, 2616G/A, 2617G/A, 2618G/A, 2619G/A, 2620G/A, 2621G/A, 2622G/A, 2623G/A, 2624G/A, 2625G/A, 2626G/A, 2627G/A, 2628G/A, 2629G/A, 2630G/A, 2631G/A, 2632G/A, 2633G/A, 2634G/A, 2635G/A, 2636G/A, 2637G/A, 2638G/A, 2639G/A, 2640G/A, 2641G/A, 2642G/A, 2643G/A, 2644G/A, 2645G/A, 2646G/A, 2647G/A, 2648G/A, 2649G/A, 2650G/A, 2651G/A, 2652G/A, 2653G/A, 2654G/A, 2655G/A, 2656G/A, 2657G/A, 2658G/A, 2659G/A, 2660G/A, 2661G/A, 2662G/A, 2663G/A, 2664G/A, 2665G/A, 2666G/A, 2667G/A, 2668G/A, 2669G/A, 2670G/A, 2671G/A, 2672G/A, 2673G/A, 2674G/A, 2675G/A, 2676G/A, 2677G/A, 2678G/A, 2679G/A, 2680G/A, 2681G/A, 2682G/A, 2683G/A, 2684G/A, 2685G/A, 2686G/A, 2687G/A, 2688G/A, 2689G/A, 2690G/A, 2691G/A, 2692G/A, 2693G/A, 2694G/A, 2695G/A, 2696G/A, 2697G/A, 2698G/A, 2699G/A, 2700G/A, 2701G/A, 2702G/A, 2703G/A, 2704G/A, 2705G/A, 2706G/A, 2707G/A, 2708G/A, 2709G/A, 2710



DE	21B4/rhoptry antigen gene 5 DNA.
XX	
KM	Polymerase chain reaction; PCR; amplify; primer; detection;
KM	babesiosis; parasite; Babesia bovis; 21B4/rhoptry; antigen; gene;
KW	repeat region; immune response; vaccine; ss.
XX	
OS	Synthetic.
XX	
PN	MO9314204-A.
PD	22-JUL-1993.
XX	
PF	15-JAN-1993; 93WO-AU00012.
PR	15-JAN-1992; 92AU-0000399.
PA	(CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI	Dalrymple BP, Peters JM;
XX	
DR	WPI; 1993-243219/30.
XX	P-PsDB; AAR39901.
PT	Detecting closely linked gene copies which encode protective
PT	antigen against babesiosis - by screening babesia genomic DNA
PT	library with oligo-nucleotide probe based partial sequencing of
PT	protective antigen and identifying positive clones
XX	
PS	Claim 23; Fig 6; 55pp; English.
XX	
CC	This sequence represents the Babesia bovis 21B4/rhoptry antigen gene
CC	5. This sequence was isolated by PCR using the primers given in
CC	A047068-72. Primer 21B4.1 corresponds to part of the repeated
CC	region of 21B4/rhoptry antigen. In hybridisation assays this primer
CC	recognised two tandemly repeated regions suggesting that B. bovis
CC	contains two copies of the 21B4/rhoptry antigen gene. The two
CC	proteins encoded by the two antigen genes are identical. Primers
CC	21B4.2 and 21B4.3 flank the 21B4-509 coding region of the antigen
CC	gene. Primer 21B4.4 primes synthesis just 3' to the end of the open
CC	reading frame. The entire open reading frame was shown to encode five
CC	antigen genes. The 3' non-repetitive sequences of open reading frames
CC	1-4 are identical. Gene 5 shows sequence divergence throughout most
CC	of the open reading frame. Babesia antigen genes can be used in the
CC	production of a combined vaccine which will stimulate a greater immune
CC	response and afford broader immunity than a single antigen vaccine.
SO	
SO	Sequence 1491 BP; 479 A; 286 C; 328 G; 398 T; 0 other;
Query Match	13.0%; Score 16; DB 14; Length 1491;
Best Local Similarity	100.0%; Pred. No. 26;
Matches 16; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	24 cttaacaacctgatga 39 
Db	941 cttaacaacctgatga 956
RESULT 15	
AAC98971/C	
ID AAC98971 standard; cDNA; 2732 BP.	
XX	
AC	AAC98971;
XX	
DT	09-MAR-2001 (first entry)
XX	
DE	Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:199.
XX	
KM	Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KM	detection; diagnosis; identification; cytostatic; neuroprotective;
KM	nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KM	antiinflammatory; cardiac; gene therapy; chromosome mapping;
KM	linkage analysis; tissue identification; tissue typing; forensic;

KM neural; immune system; muscular; reproductive; gastrointestinal;  
 KM pulmonary; cardiovascular; renal; proliferative; ss.  
 XX Homo sapiens.  
 OS WO200055320-A1.  
 PN 21-SEP-2000.  
 XX 08-MAR-2000; 2000WO-US05989.  
 XX 12-MAR-1999; 99US-0124270.  
 XX (HUMAN-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Ruben SM;  
 PI WPI: 2000-579444/54.  
 XX P-PSDB: AAB54206.  
 DR  
 XX  
 XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition -  
 XX  
 XX Claim 1; Page 645-646; 1379pp; English.  
 PS  
 XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
 CC neuroprotective, neurotropic, immunomodulatory, relaxant, contraceptive,  
 CC gynaecological, cardiant and antiinflammatory activities, and can be used  
 CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.  
 CC Agonists and antagonists to the antigens can be screened for. The  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridisation probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.  
 XX  
 XX Sequence 2732 BP; 624 A; 753 C; 688 G; 659 T; 8 other;  
 SQ  
 Query Match 13.0%; Score 16; DB 21; Length 2732;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0  
 QY 101 tcttcattatcca 116  
 Db 1444 TCTTCATTATTCCA 1429  
 ||||||||||||  
 RESULT 16  
 AAC78084/c  
 ID AAC78084 standard; cDNA; 4201 BP.  
 XX AAC78084;  
 XX 08-FEB-2001 (first entry)  
 XX Human cancer associated gene sequence SFO ID NO:478.  
 DE Human, cancer associated gene; cancer antigen; detection; cancer;  
 TW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;  
 TM



KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;  
KW antiinflammatory; antihypertoid; antiatherogenic; antibacterial; cardiant;  
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
KW vasotropic; antipsoriatic; angiogenic; gene therapy; inflammation;  
KW immune disorder; hematopoietic cell disorder; autoimmune disorder;  
KW allergic reaction; graft versus host disease; organ rejection;  
KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
KW neurological disease; drug screening; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20005350-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05882.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI: 2000-587533/55.  
DR P-PSDB: AAB43875.  
XX  
PT Novel isolated nucleic acids comprising sequences encoding peptides  
PT useful for treating or diagnosing e.g. cancer -  
XX  
PS Claim 1; Page 1012-1013; 2352pp; English.  
XX  
CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;  
CC antiinflammatory; antihypertoid; antiatherogenic; antibacterial; antiviral;  
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
CC nootropic; vasotropic; antipsoriatic and angiogenic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 4201 BP; 1323 A; 637 C; 850 G; 1381 T; 10 other;

Query Match 13.0%; Score 16; DB 21; length 4201;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 aggtatgtaaatg 17  
|||||  
Db 2555 AGGTATGTGAATGTG 2540

RESULT 17  
AAH3277/c  
ID AAH33277 standard; cDNA; 4201 BP.  
XX  
AC AAH33277;  
XX  
DT 03-SEP-2001 (first entry)  
XX

DE Human colon cancer antigen encoding cDNA SEQ ID NO:333.  
XX  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US26524.  
XX  
PR 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
DR WPI: 2001-235357/24.  
DR P-PSDB: AAG73846.  
XX  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
PS Claim 1; Page 2445-2447; 9803pp; English.  
XX  
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 4201 BP; 1323 A; 637 C; 850 G; 1381 T; 10 other;

Query Match 13.0%; Score 16; DB 22; length 4201;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 aggtatgtaaatg 17  
|||||  
Db 2555 AGGTATGTGAATGTG 2540

RESULT 18  
AAC91600/c  
ID AAC91600 standard; DNA; 5308 BP.  
XX  
AC AAC91600;  
XX  
DT 16-MAR-2001 (first entry)  
XX  
DE Human angiotensinogen gene 5' region/exon 1, SEQ ID NO:1.  
XX  
KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;



KM antidiabetic; gene therapy; 5' region; exon 1; ds.  
XX  
OS Homo sapiens.  
PN WO200071751-A1.  
PP 30-NOV-2000.  
PX  
PF 16-MAY-2000; 2000MO-USI3327.  
PR 21-MAY-1999; 99US-O135423.  
PR 06-JAN-2000; 2000US-O174700.  
PA (MYRI-) MYRIAD GENETICS INC.  
PI McGrail M, Russell DL, Shattuck DM;  
DR WPI; 2001-025172/03.  
PT Novel angiotensinogen gene, mutant alleles of which causes  
PT susceptibility to insulin-dependent diabetes mellitus useful for  
PS diagnosis of predisposition to diabetes -  
Claim 1; Page 60-63; 83pp; English.

The invention relates to the human angiotensinogen (AGT) gene, some CC  
CC mutant alleles of which cause a susceptibility to insulin-dependent CC  
CC diabetes mellitus (IDDM, type I diabetes). The AGT gene is located CC  
CC on chromosome 1q42-43, a region linked to IDDM. The invention CC  
CC discloses genomic sequences comprising exons 1-5 of the human AGT gene CC  
CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT CC  
CC gene exon 1 (AAC91605). The invention also encompasses the specifically CC  
CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the CC  
CC mutant angiotensinogen proteins ABA48945-B48949. The invention also CC  
CC relates to detecting mutant AGT alleles or gene products thereof which CC  
CC are related to IDDM; determining whether a person has, or is at risk of CC  
CC developing diabetes via detection of a polymorphism in the AGT gene; and CC  
CC methods of screening for drug candidates which may be useful in the CC  
CC treatment of diabetes resulting from an AGT mutation. Methods of CC  
CC preventing or treating diabetes are claimed which comprise the CC  
CC administration of a compound which agonises or antagonises wild-type or CC  
CC mutant AGT, which agonises or antagonises an AGT receptor which CC  
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition, CC  
CC the invention encompasses a transgenic non-human animal, or cell line CC  
CC derived therefrom, comprising a mutant human AGT allele. The CC  
CC polymorphisms identified in the AGT gene are useful for determining if a CC  
CC person has, or is at risk from developing insulin-dependent diabetes CC  
CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant CC  
CC AGT proteins or fragments thereof are useful for screening compounds CC  
CC which bind to AGT polypeptides. The present sequence represents a CC  
CC fragment of the human AGT gene comprising part of the 5' region and exon CC  
CC 1.  
Sequence 5308 BP; 1364 A; 1265 C; 1329 G; 1350 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 5308;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 71 tggtaagtcttcacatg 86  
||| ||||||| ||||| |||||  
Db 2311 TGGTAAAGTTTCACATG 2296

RESULT 19  
AAC91670/c  
xx AAC91670 standard; DNA; 5308 BP.  
AC  
xx AAC91670;  
DT 16-MAR-2001 (first entry)  
xx

DE Human angiotensinogen gene 5' region mutant, C2201T.  
XX  
XX Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
XX type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
XX IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;  
XX antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.  
XX  
XX Homo sapiens.  
CS  
XX WC200071751-A1.  
PN  
XX 30-NOV-2000.  
PD  
XX 16-MAY-2000; 2000MO-US13327.  
PE  
XX 21-MAY-1999; 990S-0135423.  
PR  
XX 06-JAN-2000; 2000US-0174700.  
XX  
XX (MYRI-) MYRIAD GENETICS INC.  
EA  
XX McGrall M, Russell DL, Shattuck DM;  
PI  
XX WPI; 2001-025172/03.  
DR  
XX  
XX Novel angiotensinogen gene, mutant alleles of which causes  
PT susceptibility to insulin-dependent diabetes mellitus useful for  
PT diagnosis of predisposition to diabetes -  
XX  
XX  
PS Claim 2; Page -: 83pp: English.  
XX  
XX The invention relates to the human angiotensinogen (AGT) gene, some  
CC mutant alleles of which cause a susceptibility to insulin-dependent  
CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located  
CC on chromosome 1q42-43, a region linked to IDDM. The invention  
CC discloses genomic sequences comprising exons 1-5 of the human AGT gene  
CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT  
CC gene exon 1 (AAC91606). The invention also encompasses the specifically  
CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the  
CC mutant angiotensinogen proteins AAB48945-B48949. The invention also  
CC relates to detecting mutant AGT alleles or gene products thereof which  
CC are related to IDDM; determining whether a person has, or is at risk of  
CC developing diabetes via detection of a polymorphism in the AGT gene; and  
CC methods of screening for drug candidates which may be useful in the  
CC treatment of diabetes resulting from an AGT mutation. Methods of  
CC preventing or treating diabetes are claimed which comprise the  
CC administration of a compound which agonises or antagonises wild-type or  
CC mutant AGT, which agonises or antagonises an AGT receptor, which  
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,  
CC the invention encompasses a transgenic non-human animal, or cell line  
CC derived therefrom, comprising a mutant human AGT allele. The  
CC polymorphisms identified in the AGT gene are useful for determining if a  
CC person has, or is at risk from developing insulin-dependent diabetes  
CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant  
CC AGT proteins or fragments thereof are useful for screening compounds  
CC which bind to AGT polypeptides. The present sequence represents a portion  
CC of the human AGT gene comprising a mutant 5' region fragment, and exon  
CC 1. Note: The present sequence is not shown in the specification, but is  
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence  
CC shown on page 60-63.  
XX  
XX  
SQ Sequence 5308 BP; 1364 A; 1264 C; 1329 G; 1351 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 5308;  
Best Local Similarity 100.0%; Prid. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Yy 71 ttgtaagttcacatg 86  
|||||  
Db 2311 TGGTAACTTTCACATG 2296

RESULT 20



AAC91671/c  
ID AAC91671 standard; DNA: 5308 BP.  
XX  
AC AAC91671:  
XX  
DT 16-MAR-2001 (first entry)  
XX  
DE Human angiotensinogen gene 5' region mutant, T2271C.  
XX  
XX Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;  
KW antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200071751-A1.  
XX  
PD 30-NOV-2000.  
XX  
PF 16-MAY-2000; 2000WO-US13327.  
XX  
PR 21-MAY-1999; 99US-0135423.  
XX  
PR 06-JAN-2000; 2000US-0174700.  
XX  
PA (MYRI-) MYRIAD GENETICS INC.  
XX  
PI McGrail M, Russell DL, Shattuck DM;  
XX  
DR WPI; 2001-025172/03.  
XX  
PT Novel angiotensinogen gene, mutant alleles of which causes  
XX susceptibility to insulin-dependent diabetes mellitus useful for  
XX diagnosis of predisposition to diabetes -  
XX  
PS Claim 2; Page -: 83pp; English.  
XX  
XX The invention relates to the human angiotensinogen (AGT) gene, some  
CC mutant alleles of which cause a susceptibility to insulin-dependent  
CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located  
CC on chromosome 1q42-43, a region linked to IDDM. The invention  
CC discloses genomic sequences comprising exons 1-5 of the human AGT gene  
CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT  
CC gene exon 1 (AAC91606). The invention also encompasses the specifically  
CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the  
CC mutant angiotensinogen proteins A848945-B48949. The invention also  
CC relates to detecting mutant AGT alleles or gene products thereof which  
CC are related to IDDM; determining whether a person has, or is at risk of  
CC developing diabetes via detection of a polymorphism in the AGT gene; and  
CC methods of screening for drug candidates which may be useful in the  
CC treatment of diabetes resulting from an AGT mutation. Methods of  
CC preventing or treating diabetes are claimed which comprise the  
CC administration of a compound which agonises or antagonises wild-type or  
CC mutant AGT, which agonises or antagonises an AGT receptor, which  
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,  
CC the invention encompasses a transgenic non-human animal, or cell line  
CC derived therefrom, comprising a mutant human AGT allele. The  
CC polymorphisms identified in the AGT gene are useful for determining if a  
CC person has, or is at risk from developing insulin-dependent diabetes  
CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant  
CC AGT proteins or fragments thereof are useful for screening compounds  
CC which bind to AGT polypeptides. The present sequence represents a portion  
CC of the human AGT gene comprising a mutant 5' region fragment, and exon  
CC 1. Note: The present sequence is not shown in the specification, but is  
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence  
XX shown on page 60-63.  
XX  
SO Sequence 5308 BP; 1364 A; 1266 C; 1329 G; 1349 T; 0 other;

QY 71 tggtgaagttcacatg 86  
|||||  
Db 2311 TGGTAACTTTCACATG 2296  
  
RESULT 21  
AAC91672/c  
ID AAC91672 standard; DNA: 5308 BP.  
XX  
AC AAC91672:  
XX  
DT 16-MAR-2001 (first entry)  
XX  
DE Human angiotensinogen gene 5' region mutant, C2731T.  
XX  
XX Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;  
KW antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200071751-A1.  
XX  
PD 30-NOV-2000.  
XX  
PF 16-MAY-2000; 2000WO-US13327.  
XX  
PR 21-MAY-1999; 99US-0135423.  
XX  
PR 06-JAN-2000; 2000US-0174700.  
XX  
PA (MYRI-) MYRIAD GENETICS INC.  
XX  
PI McGrail M, Russell DL, Shattuck DM;  
XX  
DR WPI; 2001-025172/03.  
XX  
PT Novel angiotensinogen gene, mutant alleles of which causes  
XX susceptibility to insulin-dependent diabetes mellitus useful for  
XX diagnosis of predisposition to diabetes -  
XX  
PS Claim 2; Page -: 83pp; English.  
XX  
XX The invention relates to the human angiotensinogen (AGT) gene, some  
CC mutant alleles of which cause a susceptibility to insulin-dependent  
CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located  
CC on chromosome 1q42-43, a region linked to IDDM. The invention  
CC discloses genomic sequences comprising exons 1-5 of the human AGT gene  
CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT  
CC gene exon 1 (AAC91606). The invention also encompasses the specifically  
CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the  
CC mutant angiotensinogen proteins A848945-B48949. The invention also  
CC relates to detecting mutant AGT alleles or gene products thereof which  
CC are related to IDDM; determining whether a person has, or is at risk of  
CC developing diabetes via detection of a polymorphism in the AGT gene; and  
CC methods of screening for drug candidates which may be useful in the  
CC treatment of diabetes resulting from an AGT mutation. Methods of  
CC preventing or treating diabetes are claimed which comprise the  
CC administration of a compound which agonises or antagonises wild-type or  
CC mutant AGT, which agonises or antagonises an AGT receptor, which  
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,  
CC the invention encompasses a transgenic non-human animal, or cell line  
CC derived therefrom, comprising a mutant human AGT allele. The  
CC polymorphisms identified in the AGT gene are useful for determining if a  
CC person has, or is at risk from developing insulin-dependent diabetes  
CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant  
CC AGT proteins or fragments thereof are useful for screening compounds  
CC which bind to AGT polypeptides. The present sequence represents a portion  
CC of the human AGT gene comprising a mutant 5' region fragment, and exon  
CC 1. Note: The present sequence is not shown in the specification, but is  
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence  
XX shown on page 60-63.

Query Match 13.0%; Score 16; DB 22; Length 5308;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



XX Sequence 5308 BP; 1364 A; 1264 C; 1329 G; 1351 T; 0 other;  
SQ

Query Match 13.0%; Score 16; DB 22; Length 5308;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 tggtaagttcacatg 86  
|||||  
Db 2311 TGGTAAGTTCAACATG 2296

RESULT 22  
AAC91674/C  
ID AAC91674 standard; DNA; 5308 BP.  
XX  
AC AAC91674;  
XX  
DT 16-MAR-2001 (first entry)  
XX  
DE Human angiotensinogen gene 5' region mutant, T2649A.  
XX  
KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;  
KW antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200071751-A1.  
XX  
PD 30-NOV-2000.  
XX  
PE 16-MAY-2000; 2000WO-US13327.  
XX  
PR 21-MAY-1999; 9905-0135423.  
PR 06-JAN-2000; 2000US-0174700.  
XX  
PA (MYRI-) MYRIAD GENETICS INC.  
XX  
PI McGrail M, Russell DL, Shattuck DM;  
DR WPI: 2001-025172/03.  
XX  
PT Novel angiotensinogen gene, mutant alleles of which causes  
PT susceptibility to insulin-dependent diabetes mellitus useful for  
PT diagnosis of predisposition to diabetes -  
XX  
XX Claim 2; Page -: 83pp; English.

The invention relates to the human angiotensinogen (AGT) gene, some  
CC mutant alleles of which cause a susceptibility to insulin-dependent  
CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located  
CC on chromosome 1q42-43, a region linked to IDDM. The invention  
CC discloses genomic sequences comprising exons 1-5 of the human AGT gene  
CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT  
CC gene exon 1 (AAC91606). The invention also encompasses the specifically  
CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the  
CC mutant angiotensinogen proteins AAB48945-B48949. The invention also  
CC relates to detecting mutant AGT alleles or gene products thereof which  
CC are related to IDDM; determining whether a person has, or is at risk of  
CC developing diabetes via detection of a polymorphism in the AGT gene; and  
CC methods of screening for drug candidates which may be useful in the  
CC treatment of diabetes resulting from an AGT mutation. Methods of  
CC preventing or treating diabetes are claimed which comprise the  
CC administration of a compound which agonises or antagonises wild-type or  
CC mutant AGT, which agonises or antagonises an AGT receptor, which  
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,  
CC the invention encompasses a transgenic non-human animal, or cell line  
CC derived therefrom, comprising a mutant human AGT allele. The  
CC polymorphisms identified in the AGT gene are useful for determining if a  
CC person has, or is at risk from developing insulin-dependent diabetes

CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant  
CC AGT proteins or fragments thereof are useful for screening compounds  
CC which bind to AGT polypeptides. The present sequence represents a portion  
CC of the human AGT gene comprising a mutant 5' region fragment, and exon  
CC 1. Note: The present sequence is not shown in the specification, but is  
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence  
CC shown on page 60-63.  
XX  
SQ Sequence 5308 BP; 1365 A; 1265 C; 1329 G; 1349 T; 0 other;  
XX

Query Match 13.0%; Score 16; DB 22; Length 5308;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 tggtaagttcacatg 86  
|||||  
Db 2311 TGGTAAGTTCAACATG 2296

RESULT 23  
AAC91675/C  
ID AAC91675 standard; DNA; 5308 BP.  
XX  
AC AAC91675;  
XX  
DT 16-MAR-2001 (first entry)  
XX  
DE Human angiotensinogen gene 5' region mutant, G2753A.  
XX  
KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;  
KW antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200071751-A1.  
XX  
PD 30-NOV-2000.  
XX  
PE 16-MAY-2000; 2000WO-US13327.  
XX  
PR 21-MAY-1999; 9905-0135423.  
PR 06-JAN-2000; 2000US-0174700.  
XX  
PA (MYRI-) MYRIAD GENETICS INC.  
XX  
PI McGrail M, Russell DL, Shattuck DM;  
DR WPI: 2001-025172/03.  
XX  
PT Novel angiotensinogen gene, mutant alleles of which causes  
PT susceptibility to insulin-dependent diabetes mellitus useful for  
PT diagnosis of predisposition to diabetes -  
XX  
XX Claim 2; Page -: 83pp; English.

The invention relates to the human angiotensinogen (AGT) gene, some  
CC mutant alleles of which cause a susceptibility to insulin-dependent  
CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located  
CC on chromosome 1q42-43, a region linked to IDDM. The invention  
CC discloses genomic sequences comprising exons 1-5 of the human AGT gene  
CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT  
CC gene exon 1 (AAC91606). The invention also encompasses the specifically  
CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the  
CC mutant angiotensinogen proteins AAB48945-B48949. The invention also  
CC relates to detecting mutant AGT alleles or gene products thereof which  
CC are related to IDDM; determining whether a person has, or is at risk of  
CC developing diabetes via detection of a polymorphism in the AGT gene; and  
CC methods of screening for drug candidates which may be useful in the  
CC treatment of diabetes resulting from an AGT mutation. Methods of  
CC preventing or treating diabetes are claimed which comprise the



CC administration of a compound which agonises or antagonises wild-type or  
CC mutant AGT, which agonises or antagonises an AGT receptor, which  
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,  
CC the invention encompasses a transgenic non-human animal, or cell line  
CC derived therefrom, comprising a mutant human AGT allele. The  
CC polymorphisms identified in the AGT gene are useful for determining if a  
CC person has, or is at risk from developing insulin-dependent diabetes  
CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant  
CC AGT proteins or fragments thereof are useful for screening compounds  
CC which bind to AGT polypeptides. The present sequence represents a portion  
CC of the human AGT gene comprising a mutant 5' region fragment, and exon  
CC 1. Note: The present sequence is not shown in the specification, but is  
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence  
CC shown on page 60-63.

SO Sequence 5308 BP; 1365 A; 1265 C; 1328 G; 1350 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 5308;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 tggtaagttcacatg 86  
|||||  
Db 2311 TGCTAGTTTCACATG 2296

RESULT 24  
AAC91676/c  
ID AAC91676 standard; DNA; 5308 BP.  
XX AAC91676;  
XX  
DT 16-MAR-2001 (first entry)  
XX  
XX Human angiotensinogen gene 5' region mutant, T2759C.  
DE  
XX  
XX Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
KM type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
KM IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;  
XX antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
PN WO200071751-A1.  
XX  
XX 30-NOV-2000.  
PD  
XX  
XX 16-MAY-2000; 2000WO-US13327.  
PF  
XX  
XX 21-MAY-1999; 99US-0135423.  
PR 06-JAN-2000; 2000US-0174700.  
XX  
XX (MYRI-) MYRIAD GENETICS INC.  
PA  
XX  
XX McGrail M, Russell DL, Shattuck DM;  
PI  
XX WPI; 2001-025172/03.  
DR  
XX  
XX Novel angiotensinogen gene, mutant alleles of which causes  
PT susceptibility to insulin-dependent diabetes mellitus useful for  
PT diagnosis of predisposition to diabetes -  
XX  
XX Claim 2; Page -: 83pp; English.

XX The invention relates to the human angiotensinogen (AGT) gene, some  
XX mutant alleles of which cause a susceptibility to insulin-dependent  
XX diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located  
XX on chromosome 1q42-43, a region linked to IDDM. The invention  
XX discloses genomic sequences comprising exons 1-5 of the human AGT gene  
XX (AAC91600-C91604) and a genomic sequence comprising an alternative AGT  
XX gene exon 1 (AAC91606). The invention also encompasses the specifically  
XX claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the

CC mutant angiotensinogen proteins AAB48945-B48949. The invention also  
CC relates to detecting mutant AGT alleles or gene products thereof which  
CC are related to IDDM; determining whether a person has, or is at risk of  
CC developing diabetes via detection of a polymorphism in the AGT gene; and  
CC methods of screening drug candidates which may be useful in the  
CC treatment of diabetes resulting from an AGT mutation. Methods of  
CC preventing or treating diabetes are claimed which comprise the  
CC administration of a compound which agonises or antagonises wild-type or  
CC mutant AGT, which agonises or antagonises an AGT receptor, which  
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,  
CC the invention encompasses a transgenic non-human animal, or cell line  
CC derived therefrom, comprising a mutant human AGT allele. The  
CC polymorphisms identified in the AGT gene are useful for determining if a  
CC person has, or is at risk from developing insulin-dependent diabetes  
CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant  
CC AGT proteins or fragments thereof are useful for screening compounds  
CC which bind to AGT polypeptides. The present sequence represents a portion  
CC of the human AGT gene comprising a mutant 5' region fragment, and exon  
CC 1. Note: The present sequence is not shown in the specification, but is  
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence  
CC shown on page 60-63.

SO Sequence 5308 BP; 1364 A; 1266 C; 1329 G; 1349 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 5308;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 tggtaagttcacatg 86  
|||||  
Db 2311 TGCTAGTTTCACATG 2296

RESULT 25  
AAC91677/c  
ID AAC91677 standard; DNA; 5308 BP.  
XX AAC91677;  
XX  
DT 16-MAR-2001 (first entry)  
XX  
XX Human angiotensinogen gene 5' region mutant, G2829A.  
DE  
XX  
XX Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
KM type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
KM IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;  
XX antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
PN WO200071751-A1.  
XX  
XX 30-NOV-2000.  
PD  
XX  
XX 16-MAY-2000; 2000WO-US13327.  
PF  
XX  
XX 21-MAY-1999; 99US-0135423.  
PR 06-JAN-2000; 2000US-0174700.  
XX  
XX (MYRI-) MYRIAD GENETICS INC.  
PA  
XX  
XX McGrail M, Russell DL, Shattuck DM;  
PI  
XX WPI; 2001-025172/03.  
DR  
XX  
XX Novel angiotensinogen gene, mutant alleles of which causes  
PT susceptibility to insulin-dependent diabetes mellitus useful for  
PT diagnosis of predisposition to diabetes -  
XX  
XX Claim 2; Page -: 83pp; English.

XX The invention relates to the human angiotensinogen (AGT) gene, some



CC mutant alleles of which cause a susceptibility to insulin-dependent  
 CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located  
 CC on Chromosome 1q42-43, a region linked to IDDM. The invention  
 CC discloses genomic sequences comprising exons 1-5 of the human AGT gene  
 CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT  
 CC gene exon 1 (AAC91606). The invention also encompasses the specifically  
 CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the  
 CC mutant angiotensinogen proteins AAB48945-B48949. The invention also  
 CC relates to detecting mutant AGT alleles or gene products thereof which  
 CC are related to IDDM; determining whether a person has, or is at risk of  
 CC developing diabetes via detection of a polymorphism in the AGT gene; and  
 CC methods of screening for drug candidates which may be useful in the  
 CC treatment of diabetes resulting from an AGT mutation. Methods of  
 CC preventing or treating diabetes are claimed which comprise the  
 CC administration of a compound which agonises or antagonises wild-type or  
 CC mutant AGT, which agonises or antagonises an AGT receptor, which  
 CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,  
 CC the invention encompasses a transgenic non-human animal, or cell line  
 CC derived therefrom, comprising a mutant human AGT allele. The  
 CC polymorphisms identified in the AGT gene are useful for determining if a  
 CC person has, or is at risk from developing insulin-dependent diabetes  
 CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant  
 CC AGT proteins or fragments thereof are useful for screening compounds  
 CC which bind to AGT polypeptides. The present sequence represents a portion  
 CC of the human AGT gene comprising a mutant 5' region fragment, and exon  
 CC 1. Note: The present sequence is not shown in the specification, but is  
 CC derived from the wild-type human AGT gene 5' region/exon 1 sequence  
 CC shown on page 60-63.

Sequence 5308 BP; 1365 A; 1265 C; 1328 G; 1350 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 5308;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 tggtaagttccatg 86  
 ||||||||||||||||  
 Db 2311 TGGTAAGTTTCACATG 2296

RESULT 26  
 AAX23517  
 ID AAX23517 standard; DNA; 50000 BP.  
 XX  
 AC AAX23517;  
 XX  
 DT 23-JUN-1999 (first entry)  
 XX  
 DE Human kidney aminopeptidase P genomic DNA fragment 1.  
 XX  
 KW Aminopeptidase; human; AMP; gene therapy; treatment; AMP-deficiency;  
 KW prenatal diagnosis; angioedema; antihypertensive agent; atherosclerosis;  
 KW arterial stenosis; industrial protein feed; malabsorption syndrome;  
 KW proteinaceous waste degradation; additive; immunohistochemistry; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9911799-A2.  
 XX  
 PD 11-MAR-1999.  
 XX  
 PF 02-SEP-1998; 98WO-US18426.  
 XX  
 PR 02-SEP-1997; 97US-0057854.  
 XX  
 XX (MEDT-) MEDICAL COLLEGE GEORGIA RES INST.  
 XX  
 PI Ryan JW, Sprinkle JJC, Venema RC;  
 XX  
 DR WPL; 1999-205193/17.  
 XX  
 PT Nucleic acid encoding human aminopeptidase P

XX  
 PS Claim 13; Page 80-109; 201pp; English.  
 XX  
 CC This invention describes the isolation of a novel human aminopeptidase P  
 CC (Amp). This protein is used to produce recombinant Amp and can be used  
 CC for gene therapy for treating AMP-deficiency conditions. Its fragments  
 CC are used as primers and probes to identify patients with homozygous and  
 CC heterozygous Amp deficiency, including prenatal diagnosis (patients  
 CC defective in Amp are at risk of developing angioedema if treated with  
 CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors  
 CC in cases of excessive Amp expression. The product of the invention is  
 CC also used to identify Amp-expressing sequences in other animals and to  
 CC generate transgenic animals, and comparisons of genomic sequences are  
 CC used to detect mutations. Amp inhibitors are potentially useful as  
 CC antihypertensive agents and to prevent or treat arterial (re)stenosis  
 CC or atherosclerosis. The structure of Amp is used to design synthetic  
 CC substrates, e.g. for use in Amp assays. Amp, which hydrolyzes N-terminal  
 CC imido bonds, can be used to degrade industrial protein feeds to free  
 CC amino acids, to degrade proteinaceous wastes, as additives in enzyme  
 CC formulations used to treat malabsorption syndrome and for studying its  
 CC biological role. Antibodies against Amp are used in immunohistochemical  
 CC methods to study Amp distribution.

Sequence 50000 BP; 13187 A; 12125 C; 11842 G; 12846 T; 0 other;

Query Match 13.0%; Score 16; DB 20; Length 50000;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 acaacctgtatgacct 42  
 ||||||||||||||||  
 Db 47023 acaacctgtatgacct 47038

RESULT 27  
 AAC91627/C  
 ID AAC91627 standard; DNA; 20 BP.  
 XX  
 AC AAC91627;  
 XX  
 DT 16-MAR-2001 (first entry)  
 XX  
 DE Human angiotensinogen gene promoter/5' genomic PCR primer, SEQ ID NO:29.  
 XX  
 KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;  
 KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;  
 KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;  
 KW antidiabetic; gene therapy; 5' region; promoter; PCR primer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200071751-A1.  
 XX  
 PD 30-NOV-2000.  
 XX  
 PF 16-MAY-2000; 2000WO-US13327.  
 XX  
 PR 21-MAY-1999; 99US-0135423.  
 XX  
 PR 06-JAN-2000; 2000US-0174700.  
 XX  
 PA (MYRIAD GENETICS INC.  
 XX  
 PI McGrail M, Russell DL, Shattuck DM;  
 XX  
 DR WPL; 2001-025172/03.  
 XX  
 XX Novel angiotensinogen gene, mutant alleles of which causes  
 XX susceptibility to insulin-dependent diabetes mellitus useful for  
 XX diagnosis of predisposition to diabetes -  
 XX  
 PS Example 2; Page 33; 83pp; English.



CC The invention relates to the human angiotensinogen (AGT) gene, some  
 CC mutant alleles of which cause a susceptibility to insulin-dependent  
 CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located  
 CC on chromosome 1q42-43, a region linked to IDDM. The invention  
 CC discloses genomic sequences comprising exons 1-5 of the human AGT gene  
 CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT  
 CC gene exon 1 (AAC91605). The invention also encompasses the specifically  
 CC claimed human AGT mutant nucleic acid sequences AAC91607-C91684, and the  
 CC mutant angiotensinogen proteins AAB48945-B48949. The invention also  
 CC relates to detecting mutant AGT alleles or gene products thereof which  
 CC are related to IDDM, determining whether a person has, or is at risk of  
 CC developing diabetes via detection of a polymorphism in the AGT gene; and  
 CC methods of screening for drug candidates which may be useful in the  
 CC treatment of diabetes resulting from an AGT mutation. Methods of  
 CC preventing or treating diabetes are claimed which comprise the  
 CC administration of a compound which agonises or antagonises wild-type or  
 CC mutant AGT, which agonises or antagonises an AGT receptor, which  
 CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,  
 CC the invention encompasses a transgenic non-human animal, or cell line  
 CC derived therefrom, comprising a mutant human AGT allele. The  
 CC polymorphisms identified in the AGT gene are useful for determining if a  
 CC person has, or is at risk from developing insulin-dependent diabetes  
 CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant  
 CC AGT proteins or fragments thereof are useful for screening compounds  
 CC which bind to AGT polypeptides. The present sequence represents a human  
 CC AGT gene 5' region/promoter PCR primer used in an exemplification of the  
 CC invention.

XX  
 SQ Sequence 20 BP; 5 A; 5 C; 4 G; 6 T; 0 other;

Query Match 12.2%; Score 15; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 ggttaagttcacatg 86  
 |||||||||||||  
 Db 20 GGTAAGTTTCACATG 6

#### RESULT 28

AAV72858  
 ID AAV72858 standard; DNA; 96 BP.

AC AAV72858;

DT 22-FEB-1999 (first entry)

DE Tryptophan 2,3-dioxygenase oligonucleotide variant.

XX  
 XX  
 KM Human: tryptophan 2,3-dioxygenase; TDO2; allelic polymorphisms;  
 KM reward deficiency syndrome; RDS; obesity; alcohol; tobacco; drug use;  
 KM Tourette syndrome; attention deficit hyperactivity disorder; ADHD;  
 KM schizoid/avoidant behaviour; aggression; premenstrual syndrome;  
 KM violence; hostility; mania; depression; anxiety; sleep problem; autism;  
 KM osteoporosis; binge eating; craving; inhibition; PCR primer; ss.

OS Synthetic.

OS Homo sapiens.

PN WO9848785-A2.

PD 05-NOV-1998.

PF 29-APR-1998; 98WO-US08684.

PR 29-APR-1997; 97US-0044394.

XX  
 XX  
 PA (BLUM-) BLUM INC KENNETH.  
 PA (CITY ) CITY OF HOPE NAT MEDICAL CENT.  
 PA (TEXA ) UNIV TEXAS SYSTEM.

XX Blum K, Comings DE, Ivy JL;

XX  
 DR WPI; 1998-610008/51.  
 XX  
 PT Composition for treating reward deficiency syndrome behaviour,  
 PT attention deficit disorders or controlling weight - contains  
 PT inhibitor of opiate peptide destruction, neurotransmitter precursor  
 PT and chromium

PS Example 10; Page 293; 663pp; English.

XX  
 CC A composition (A) has been developed for: (i) treating a reward  
 CC deficiency syndrome (RDS) behaviour; or (ii) preventing or treating  
 CC unwanted weight gain. (A) comprises: (a) an agent (I), i.e. amino acid  
 CC or peptide, or their analogues or derivatives, that inhibits enzymatic  
 CC destruction of a neuropeptidyl opiate; (b) a neurotransmitter precursor  
 CC (II) to promote neurotransmitter synthesis, i.e. L-tyr, L-phe or L-dopa  
 CC (dopamine precursors) L-tyr or 5-hydroxytryptophan (serotonin  
 CC precursors); and (C) chromium picolinate or nicotinate to increase the  
 CC level of Trp. Typical of many behaviours that can be treated with (A)  
 CC include: substance use disorders; obesity; alcohol, tobacco or other  
 CC drug use; Tourette syndrome; attention deficit hyperactivity disorder  
 CC (ADHD); schizoid/avoidant behaviour; aggression; premenstrual syndrome;  
 CC violence; hostility; mania; depression; anxiety; sleep problems; autism.  
 CC Methods given in the invention can be applied for diagnosis of RDS or  
 CC e.g. elevated levels of low density lipoprotein (LDL) or cholesterol,  
 CC longevity, lack of ADHD, osteoporosis. In treatment of obesity, (A)  
 CC inhibits binge eating and craving. The present sequence represents an  
 CC oligonucleotide used in an example from the present invention.

SQ Sequence 96 BP; 38 A; 7 C; 20 G; 31 T; 0 other;

Query Match 12.2%; Score 15; DB 19; Length 96;

Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtatcgttaaatgtg 17  
 |||||||||||||  
 Db 15 ggtatcgttaaatgtg 29

#### RESULT 29

AAC65347  
 ID AAC65347 standard; DNA; 96 BP.

AC AAC65347;

DT 12-FEB-2001 (first entry)

DE Human TDO2 gene G to T variant oligomer #1.

XX  
 XX  
 KM Human: anti-ADD; anti-ADHD; opiate destruction inhibitor;  
 KM neurotransmitter synthesis promoter; tryptophan-concentration enhancing;  
 KM allelic polymorphisms; reward deficiency syndrome; RDS; obesity;  
 KM smoking; Tourette's syndrome; schizoid avoidant behaviour; aggression;  
 KM post-traumatic stress syndrome; pre-menstrual syndrome;  
 KM cannabinoid receptor gene; dopamine DMD4 receptor;  
 KM tryptophan 2,3-dioxygenase; TDO2; PCR primer; ss.

OS Homo sapiens.

PN US6132724-A.

PD 17-OCT-2000.

PF 29-APR-1998; 98US-0069886.

PR 29-APR-1998; 98US-0069886.

XX  
 XX  
 PA (CITY ) CITY OF HOPE NAT MEDICAL CENT.  
 PA (TEXA ) UNIV TEXAS SYSTEM.







PT have of neuropeptidyl opiate enzymatic destruction inhibitor,  
PT neurotransmitter, mineral compound and Rhodiola extract or huperzine -  
XX  
XX Example 19; Column 233; 207pp; English.  
XX  
CC The present sequence is used in a method for the allelic polygene  
CC diagnosis of reward deficiency syndrome. Compositions are disclosed which  
CC comprise at least one substance that inhibits the enzymatic destruction  
CC of a neuropeptidyl opiate, at least one neurotransmitter, a tryptophan  
CC concentration-enhancing amount of a mineral compound; and at least one  
CC substance chosen from Rhodiola extract and huperzine. The compositions  
CC are used to treat ADD and ADHD. They are used to treat reward deficiency  
CC syndrome (RDS) behaviours including obesity, smoking, Tourette's  
CC syndrome, schizoid avoidant behaviour, aggression, post-traumatic stress  
CC syndrome, pre-menstrual syndrome or tobacco use. The presence of  
CC enkephalin releasers dramatically improves the patient's response to  
CC treatment.  
XX  
XX Sequence 101 BP; 40 A; 7 C; 22 G; 32 T; 0 other;  
SQ

Query Match 12.2%; Score 15; DB 21; Length 101;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggcattgtaaatgt 17  
|||||  
Db 15 ggcattgtaaatgt 29

RESULT 32  
AAZ36299  
ID AAZ36299 standard; cDNA; 201 BP.  
XX  
AC AAZ36299;  
XX  
DT 22-FEB-2000 (first entry)  
XX  
DE Mechanical stress induced glutamyl-cystein synthetase cDNA.  
XX  
KW Expressed sequence tag; EST; mechanical stress; gene therapy;  
KM osteoporosis; bone density; bone development; ss.  
XX  
OS Rattus sp.  
XX  
PN WO960164-A1.  
XX  
PD 25-NOV-1999.  
XX  
PF 14-MAY-1999; 99WO-US11066.  
XX  
PR 15-MAY-1998; 98US-0085673.  
XX  
PA (OVAR-) QDARK BIOTECH INC.  
XX  
PI Elmat P, Mor O, Skallier R, Feinstein E, Faerman A;  
XX  
XX WPI; 2000-053304/04.  
XX  
PT Identification of stress induced genes for determining risk and  
PT preventing, treating or controlling osteoporosis -  
XX  
XX Claim 25; Fig 2A-5; 308pp; English.  
XX  
CC AAZ36299-236320 represent genes or expressed sequence tags (ESTs)  
CC identified by the method of the invention. The sequences were isolated  
CC from rat osteoblasts. The specification describes a method for the  
CC identification of genes responsive to a specific mechanical stress. The  
CC method comprises applying the mechanical stress to an organism (tissue  
CC or cells comprising bone cells), isolating the specific cellular  
CC fractions and extracting mRNA from them, and differentially analysing the  
CC mRNA in comparison with control samples. The method is used to identify  
CC genes whose expression is responsive to a specific stress. The identified

CC genes are employed in determining risk associated with a physiological  
CC or disease state. The risk determination methods are used for testing a  
CC medicament for gene therapy. These medicaments, or genes identified by  
CC the method of the invention, are used for treating, preventing or  
CC controlling a physiological or disease state (especially osteoporosis  
CC or bone density or other factors causing or contributing to osteoporosis  
CC or its symptoms or other conditions involved in mechanical stress or its  
CC lack. The methods can also be used for advancing research or studies in  
CC bone development.  
XX  
XX Sequence 201 BP; 56 A; 42 C; 50 G; 53 T; 0 other;  
SQ

Query Match 12.2%; Score 15; DB 21; Length 201;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 aggtattgtaaatgt 16  
|||||  
Db 121 aggtattgtaaatgt 135

RESULT 33  
AAC13255/C  
ID AAC13255 standard; cDNA; 203 BP.  
XX  
AC AAC13255;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein 5' EST, SEQ ID NO: 17330.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KM gene therapy; chromosome mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PA (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI; 2000-500381/45.  
XX  
DR  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 17330; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
XX Sequence 203 BP; 57 A; 44 C; 35 G; 59 T; 8 other;



Query Match 12.2%; Score 15; DB 21; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 gttcacatgtaag 91  
 Db 15 GTTTCACATGTAAG 1

## RESULT 34

AAA01492  
 ID AAA01492 standard; cDNA; 280 BP.

AC AAA01492;

DT 19-MAY-2000 (first entry)

DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1483.

XX Human; colon cancer; tumour; diagnosis; gene expression product;

KM probe; detection; cancerous state; metastasis; identification;

KW breast cancer; oestrogen receptor-positive breast cancer; therapy;

XX Oestrogen receptor-negative breast cancer; lung cancer; ss.

OS Homo sapiens.

PN MO9958675-A2.

PD 18-NOV-1999.

PE 13-MAY-1999; 99WO-US10602.

PF 14-MAY-1998; 98US-0085426.

PR 15-MAY-1998; 98US-0085537.

PR 15-MAY-1998; 98US-0085696.

PR 21-OCT-1998; 98US-0105234.

PR 27-OCT-1998; 98US-0105877.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PT Polynucleotide library used to determine cancerous states of mammalian

PS cells - Claim 1; Page 537; 1097pp; English.

XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA

XX libraries constructed from human colon cancer cell lines. The present

XX invention also describes a method of detecting differentially expressed

XX genes correlated with a cancerous state of a mammalian cell, comprising

XX detecting at least one differentially expressed gene product in a test

XX sample derived from a cell suspected of being cancerous, where detection

XX of the differentially expressed gene product is correlated with a

XX cancerous state of the cell from which the test sample was derived.

XX The polynucleotides sequences can be used in a method for detecting

XX differentially expressed genes correlated with a cancerous state of a

XX mammalian cell. The polynucleotides can also be used as probes for

XX detecting and mapping related genes. They can be used in diagnosis and

Sequence 280 BP; 63 A; 45 C; 51 G; 107 T; 14 other;

Query Match 12.2%; Score 15; DB 21; Length 280;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 tcaattattccaca 118  
 Db 83 tcaattattccaca 97

## RESULT 35

AAZ14187/C  
 ID AAZ14187 standard; cDNA; 300 BP.

XX AAZ14187;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:1656.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KM detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

PN MO9938972-A2.

PD 05-AUG-1999.

PE 28-JAN-1999; 99WO-US01619.

PF 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PT Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PS Escobedo J, Garcia A, Kennedy GC, Innis MA;

XX Jones WL, Kassam A, Garcia V, Kita D, Labat I;

XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

XX Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

XX differentially expressed in different cell types

XX Claim 1; Page 1035; 2479pp; English.

XX The present invention describes a library of human polynucleotides

XX comprising the sequences given in AAZ12532 to AAZ1779. Also described is

XX a method of detecting differentially expressed genes correlated with the

XX cancerous state of a mammalian cell, comprising detecting at least one

XX differentially expressed gene product in a test sample from a cell

Sequence 280 BP; 63 A; 45 C; 51 G; 107 T; 14 other;

Query Match 12.2%; Score 15; DB 21; Length 280;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 tcaattattccaca 118  
 Db 83 tcaattattccaca 97

## RESULT 35

AAZ14187/C  
 ID AAZ14187 standard; cDNA; 300 BP.

XX AAZ14187;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:1656.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KM detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

PN MO9938972-A2.

PD 05-AUG-1999.

PE 28-JAN-1999; 99WO-US01619.

PF 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PT Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PS Escobedo J, Garcia A, Kennedy GC, Innis MA;

XX Jones WL, Kassam A, Garcia V, Kita D, Labat I;

XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

XX Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

XX differentially expressed in different cell types

XX Claim 1; Page 1035; 2479pp; English.

XX The present invention describes a library of human polynucleotides

XX comprising the sequences given in AAZ12532 to AAZ1779. Also described is

XX a method of detecting differentially expressed genes correlated with the

XX cancerous state of a mammalian cell, comprising detecting at least one

XX differentially expressed gene product in a test sample from a cell



CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.

XX  
SQ Sequence 300 BP; 90 A; 50 C; 66 G; 94 T; 0 other;

Query Match 12.2%; Score 15; DB 20; Length 300;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 tgccttacaacctg 35  
|||||  
Db 66 TGCCTTACAACCTG 52

RESULT 36  
AAH67192 standard; DNA; 303 BP.

XX  
AC AAH67192;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum coding sequence fragment SEQ ID NO: 2227.

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KM organic acid synthesis; ds.

XX  
OS Corynebacterium glutamicum.

PN EP1108790-A2.

XX  
PD 20-JUN-2001.

XX  
PF 18-DEC-2000; 2000EP-0127688.

XX  
PR 16-DEC-1999; 99JP-0377484.

XX  
PR 07-APR-2000; 2000JP-0159162.

XX  
PR 03-AUG-2000; 2000JP-0280988.

XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX  
DR WPI; 2001-376931/40.

XX  
DR P-PSDB; AAG91973.

XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

XX  
PT  
XX  
PS Claim 8; SEQ ID NO: 2227; 246bp + Sequence Listing; English.

XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX  
SQ Sequence 303 BP; 84 A; 80 C; 69 G; 70 T; 0 other;

Query Match 12.2%; Score 15; DB 22; Length 303;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 109 ttattccacacaca 123  
|||||  
Db 137 ttattccacacaca 151

RESULT 37  
AAC74896/c  
ID AAC74896 standard; cDNA; 309 BP.

XX  
AC AAC74896;

XX  
DT 08-FEB-2001 (first entry)

XX  
DE Human ORFX ORF451 polynucleotide sequence SEQ ID NO:901.

KW Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.

XX  
OS Homo sapiens.

XX  
PN WO200058473-A2.

XX  
PD 05-OCT-2000.

XX  
PF 31-MAR-2000; 2000WO-US08621.

XX  
PR 31-MAR-1999; 99US-0127607.

XX  
PR 02-APR-1999; 99US-0127636.

XX  
PR 05-APR-1999; 99US-0127728.

XX  
PR 30-MAR-2000; 2000US-0540763.

XX  
PA (CURA-) CURAGEN CORP.

XX  
PI Shinkets RA, Leach M;

XX  
DR WPI; 2000-602362/57.

XX  
DR P-PSDB; AAB40687.

XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -

XX  
PT  
XX  
PS Claim 5; Page 916; 5507bp; English.

XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy



CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, anti-inflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

CC Sequence 309 BP; 89 A; 81 C; 59 G; 80 T; 0 other;

CC Query Match

Best Local Similarity 12.2%; Score 15; DB 21; Length 309;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC Db 8 ttttaattgtgagatg 22

CC 224 TGTAAATGTGAGATG 210

CC RESULT 38

CC AAX40689/C

CC ID AAX40689 standard; cDNA; 319 BP.

CC AC AAX40689;

CC DT 18-JUN-1999 (first entry)

CC DE Human secreted protein 5' EST SEQ ID NO: 289.

CC KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 CC forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
 CC upstream regulatory sequence; cytokine activity; cell proliferation;  
 CC differentiation; haematopoiesis regulation; tissue growth regulation;  
 CC reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 CC thrombolytic; anti-inflammatory; tumour inhibition; ds.

CC OS Homo sapiens.

CC PN W09906550-A2.

CC PD 11-FEB-1999.

CC PF 31-JUL-1998; 98WO-IB01232.

CC PR 01-AUG-1997; 97US-0905144.

CC PA (GEST ) GENSET.

CC PI Duclert A, Dumas Milne Edwards J, Lacroix B;

CC DR WPI: 1999-153780/13.

CC P-PSDB: AAY11967.

CC PT New isolated prostate-derived nucleic acids - used to develop  
 CC products which may have cytokine, immune regulatory, haematopoiesis  
 CC regulating, anti-inflammatory or tumour inhibition activity

CC XX Claim 1; Page 453; 675pp; English.

CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins expressed in prostate, and encode the proteins  
 CC given in AAY11716 to AAY11993 respectively. The proteins given represent  
 CC the signal peptide and an N-terminal fragment of a secreted protein. The  
 CC nucleic acid sequences can be used for producing secreted human gene  
 CC products. They can also be used to develop products for diagnosis and  
 CC therapy. The proteins obtained may have cytokine activity, cell  
 CC proliferation and differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptides can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.

CC Sequence 319 BP; 61 A; 83 C; 110 G; 58 T; 7 other;

CC Query Match

Best Local Similarity 12.2%; Score 15; DB 20; Length 319;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC Db 56 gagggctgactgac 70

CC 112 GAGGCGTCGACGTGAC 98

CC RESULT 39

CC AAC2215/C

CC ID AAC22215 standard; cDNA; 357 BP.

CC AC AAC22215;

CC DT 06-OCT-2000 (first entry)

CC DE Human secreted protein 5' EST, SEQ ID NO: 26290.

CC KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 CC gene therapy; chromosome mapping; ss.

CC OS Homo sapiens.

CC PN EP103401-A2.

CC PD 06-SEP-2000.

CC PF 21-FEB-2000; 2000EP-0200610.

CC PR 26-FEB-1999; 99US-0122487.

CC PA (GEST ) GENSET.

CC PI Dumas Milne Edwards J, Duclert A, Giordano J;

CC DR WPI: 2000-500381/45.

CC PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 CC obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 CC Claim 1; SEQ ID 26290; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

CC Sequence 357 BP; 118 A; 56 C; 68 G; 111 T; 4 other;

CC Query Match 12.2%; Score 15; DB 21; Length 357;



Best Local Similarity 100.0%; Pred. No. 91;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 ttatctcaatttat 112  
|||||  
Db 202 TTATCTTCAATTAT 188

## RESULT 40

AAF66728  
ID AAF66728 standard; cDNA; 377 BP.

AC AAF66728;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 2484.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;

KM breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Williams LR, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LM, Strache-Crain B;

DR WPI; 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a  
mamalian cell and detecting cancer, particularly of the colon or  
prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 910; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human  
polynucleotides. The library is used to detect differentially expressed  
genes correlated with a cancerous state of a mamalian cell and can  
detect colon, prostate, breast and lung cancer. The library can be used  
to produce probes for detection of mRNA and to produce additional copies  
of the polynucleotides. The probes can be used for chromosome mapping of  
the polynucleotide and for detection of transcription levels. Ribozymes  
or antisense oligonucleotides can be generated. The polynucleotides and  
their gene products are used as genetic or biochemical markers (e.g. in  
blood or tissues) that will detect the earliest changes along the  
carcinogenesis pathway and/or monitor the efficacy of therapies and  
preventive interventions. The polynucleotides, polypeptides and  
antibodies against them can be used in pharmaceutical compositions to  
treat the cancers and proliferative disorders such as neoplasia,  
dysplasia and hyperplasia.

CC Sequence 377 BP; 116 A; 57 C; 76 G; 128 T; 0 other;

Query Match 12.2%; Score 15; DB 22; Length 377;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtatgttaaatgt 17  
|||||

Db 20 ggtatgttaaatgt 34

## RESULT 41

AAF66301/C  
ID AAF66301 standard; cDNA; 391 BP.

XX AAF66301;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 2057.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;

KM breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Williams LR, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LM, Strache-Crain B;

DR WPI; 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a  
mamalian cell and detecting cancer, particularly of the colon or  
prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 842; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human  
polynucleotides. The library is used to detect differentially expressed  
genes correlated with a cancerous state of a mamalian cell and can  
detect colon, prostate, breast and lung cancer. The library can be used  
to produce probes for detection of mRNA and to produce additional copies  
of the polynucleotides. The probes can be used for chromosome mapping of  
the polynucleotide and for detection of transcription levels. Ribozymes  
or antisense oligonucleotides can be generated. The polynucleotides and  
their gene products are used as genetic or biochemical markers (e.g. in  
blood or tissues) that will detect the earliest changes along the  
carcinogenesis pathway and/or monitor the efficacy of therapies and  
preventive interventions. The polynucleotides, polypeptides and  
antibodies against them can be used in pharmaceutical compositions to  
treat the cancers and proliferative disorders such as neoplasia,  
dysplasia and hyperplasia.

CC Sequence 391 BP; 85 A; 105 C; 130 G; 71 T; 0 other;

Query Match 12.2%; Score 15; DB 22; Length 391;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 gagagctgcactgac 70  
|||||

Db 105 GAGGCTGCACCTGAC 91

## RESULT 42

AAF66405/C



ID	AAAF66405 standard; cDNA; 412 BP.
AC	AAAF66405;
XX	
DT	09-APR-2001 (first entry)
XX	
DE	Novel human polynucleotide, SEQ ID NO: 2161.
XX	
KW	Human; cytostatic; gene therapy; colon cancer; prostate cancer; breast cancer; lung cancer; cancer detection; ss.
XX	
OS	Homo sapiens.
PN	WO200102568-A2.
PD	11-JAN-2001.
XX	
PF	30-JUN-2000; 2000MO-US18374.
XX	
PR	02-JUL-1999; 99US-0142310.
PR	02-JUL-1999; 99US-0142311.
XX	
PA	(CHIR ) CHIRON CORP.
PA	(HYSE-) HYSEQ INC.
P1	Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R, Cirkinjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D; Kita D, Garcia V, Jones LW, Strache-Crain B; WPI; 2001-091805/10.
PT	Library of polynucleotides for diagnosing a cancerous state of a mammalian cell and detecting cancer, particularly of the colon or prostate, comprises 3351 human polynucleotide sequences -
PS	Claim 9; Page 859; 1046pp; English.
CC	The present sequence is one of 3351 sequences in a library of human polynucleotides. The library is used to detect differentially expressed genes correlated with a cancerous state of a mammalian cell and can detect colon, prostate, breast and lung cancer. The library can be used to produce probes for detection of mRNA and to produce additional copies of the polynucleotides. The probes can be used for chromosome mapping of the polynucleotide and for detection of transcription levels. Ribozymes or antisense oligonucleotides can be generated. The polynucleotides and their gene products are used as genetic or biochemical markers (e.g. in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or monitor the efficacy of therapies and preventive interventions. The polynucleotides, polypeptides and antibodies against them can be used in pharmaceutical compositions to treat the cancers and proliferative disorders such as neoplasia, dysplasia and hyperplasia.
SO	Sequence 412 BP; 113 A; 99 C; 85 G; 115 T; 0 other;
OY	60 gctgcactgactgtgt 74       
DB	333 GCTGCACGTGACTGGT 319
Query Match	12.2%; Score 15; DB 22; Length 412;
Best Local Similarity	100.0%; Pred. No. 91;
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RESULT	43
AAC01762/c	
ID	AAC01762 standard; cDNA; 434 BP.
AC	AAC01762;
XX	
DT	06-OCT-2000 (first entry)

XX		Human secreted protein 5' EST, SEQ ID NO: 1760.
DE		
XX		
KM		Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW		gene therapy; chromosome mapping; ss.
XX		
OS		Homo sapiens.
XX		
PN		EPI033401-A2.
XX		
PD		06-SEP-2000.
XX		
PE		21-FEB-2000; 2000EP-0200610.
XX		
PR		26-FEB-1999; 99US-0122487.
XX		
PA		(GEST ) GENSET.
XX		
Dumas Milne Edwards J,		Duclet A, Giordano J;
DR		
P-PSDB: AAG01756.		
XX		
PT		New nucleic acid that is a 5' expressed sequence tag (5' EST) for
diagnostic, forensic, gene therapy and chromosome mapping procedures -		
Claim 1; SEQ ID 1760; 71bp + CD-ROM; English.		
XX		
CC		The present sequence is one of a large number of 5' ESTs derived from
CC		mRNAs encoding secreted proteins. An ORF has been identified within the
sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs		
derived from 30 different tissues. EST sequences usually correspond		
mainly to the 3' untranslated region (UTR) of the mRNA because they are		
often obtained from oligo-dT primed cDNA libraries. Such ESTs are not		
well suited for isolating cDNA sequences derived from the 5' ends of		
mRNAs and even in those cases where longer cDNA sequences have been		
obtained, the full 5' UTR is rarely included. 5' ESTs are derived from		
mRNAs with intact 5' ends and can therefore be used to obtain full length		
cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,		
gene therapy and chromosome mapping procedures. They are used to obtain		
upstream regulatory sequences and to design expression and secretion		
vectors.		
XX		
SO		Sequence 434 BP: 123 A; 116 C; 124 G; 70 T; 1 other;
Query Match	12.2%; Score 15; DB 21; Length 434;	
Best Local Similarity	100.0%; Pred. No. 91;	
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	87 taagggtacattat 101	
DB	375 TAAGTGATCTACTAT 361	
RESULT 44		
AAFI0271		
ID	AAFI0271 standard; cDNA; 556 BP.	
XX		
AC	AAFI0271:	
XX		
DT	13-MAR-2001 (first entry)	
XX		
DE	Fusarium venenatum EST SEQ ID NO:2794.	
XX		
KM	Multiple gene expression; filamentous fungal cell; EST;	
KW	expressed sequence tag; Fusarium venenatum; Aspergillus niger;	
KM	Aspergillus oryzae; Trichoderma reesei; identification; recombination;	
KW	culture condition; environmental stress; spore morphogenesis;	
KM	metabolic pathway engineering; catabolic pathway engineering; ss.	
XX		
OS	Fusarium venenatum.	



XX WO200056762-A2.  
 XX 28-SEP-2000.  
 XX 22-MAR-2000; 2000WO-US07781.  
 XX 22-MAR-1999; 99US-0273623.  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 XX (NOVO ) NOVO NORDISK AS.  
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB,  
 XX WPI; 2000-594572/56.  
 XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX  
 XX Claim 86; Page 1396; 3161pp; English.  
 XX The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AA07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 XX Sequence 556 BP; 160 A; 128 C; 138 G; 128 T; 2 other;  
 XX  
 XX Query Match 12.2%; Score 15; DB 21; Length 556;  
 XX Best Local Similarity 100.0%; Pred No. 90;  
 XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX 07-MAR-2001 (first entry)  
 XX Human CLASP-2 genomic DNA from BAC ref 13.1.  
 XX  
 XX CLASP-1; CLASP-2; transmembrane protein; immune response; inflammatory;  
 KW cadherin-like asymmetry protein; autoimmune disease; immunosuppressive;  
 KW immunomodulatory; antiinflammatory; antiarthritic; cytostatic;  
 KW hypotensive; antirheumatic; antinaemic; haemostatic; neuroprotective;

KW hypersensitivity; transplantation rejection response; immunodeficiency;  
 KW proliferation; differentiation; inflammatory response; arthritis;  
 KW inflammatory bowel disease; haematopoietic cell; blood protein disorder;  
 KW anaemia; thrombocytopenia; multiple sclerosis; rheumatoid arthritis;  
 KW endometriosis; pregnancy induced hypertension; ds.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200061747-A2.  
 XX 19-OCT-2000.  
 XX  
 XX 13-APR-2000; 2000WO-US10158.  
 XX  
 XX 14-APR-1999; 99US-0129171.  
 XX 14-MAY-1999; 99US-0134114.  
 XX 14-MAY-1999; 99US-0134117.  
 XX 14-MAY-1999; 99US-0134118.  
 XX 21-OCT-1999; 99US-0160860.  
 XX 29-OCT-1999; 99US-0162498.  
 XX 13-DEC-1999; 99US-0170453.  
 XX 14-JAN-2000; 2000US-0176195.  
 XX 14-FEB-2000; 2000US-0182296.  
 XX  
 XX (ARBO-) ARBOR VITA CORP.  
 XX  
 XX Lu PS;  
 XX  
 XX WPI; 2000-619230/59.  
 XX  
 XX Isolated cadherin-like asymmetry protein-2 polynucleotide and  
 PT polypeptide used to diagnose, treat and prevent autoimmune diseases and  
 PT inflammatory responses -  
 XX  
 XX Example 3; Fig 6A; 286pp; English.  
 XX  
 XX The present invention describes cadherin-like asymmetry protein-2  
 CC (CLASP-2). CLASP-2 can have immunosuppressive, immunomodulatory,  
 CC antiinflammatory, antiarthritic, cytostatic, hypotensive, antirheumatic,  
 CC antinaemic, haemostatic and neuroprotective activities. CLASP-2 can be  
 CC used to inhibit an immune response in a subject by interfering with the  
 CC ability of a CLASP-2 protein to bind to another T cell or B cell. An  
 CC immune response in a subject may also be inhibited by administering an  
 CC antibody which specifically binds to CLASP-2. CLASP-2 polynucleotides,  
 CC proteins and antibodies can be used to prevent or treat a CLASP-2  
 CC mediated disease, such as an autoimmune disease caused or exacerbated  
 CC by increased activity of TH1 cells. They can also be used to treat  
 CC hypersensitivities, prevent transplantation rejection responses and  
 CC augment immune responsiveness in immunodeficiency states, inhibit  
 CC proliferation and differentiation of cells involved in an inflammatory  
 CC response e.g. arthritis, inflammatory bowel disease and increase  
 CC differentiation and proliferation of haematopoietic cells e.g. to treat  
 CC anaemia, thrombocytopenia and other blood protein disorders. Disorders  
 CC treated by disrupting CLASP-2 function include multiple sclerosis,  
 CC rheumatoid arthritis, endometriosis and pregnancy induced hypertension.  
 CC The present sequence represents a human CLASP-2 genomic DNA sequence  
 CC from a bacterial artificial chromosome (BAC), given in the present  
 CC invention.  
 XX  
 XX Sequence 567 BP; 161 A; 89 C; 105 G; 188 T; 24 other;  
 XX  
 XX Query Match 12.2%; Score 15; DB 21; Length 567;  
 XX Best Local Similarity 100.0%; Pred No. 90;  
 XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX 07-MAR-2001 (first entry)  
 XX Human CLASP-2 genomic DNA from BAC ref 13.1.  
 XX  
 XX CLASP-1; CLASP-2; transmembrane protein; immune response; inflammatory;  
 KW cadherin-like asymmetry protein; autoimmune disease; immunosuppressive;  
 KW immunomodulatory; antiinflammatory; antiarthritic; cytostatic;  
 KW hypotensive; antirheumatic; antinaemic; haemostatic; neuroprotective;



Job time: 2912 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 10:00:11 ; Search time 39.53 Seconds  
(without alignments)  
704.700 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123  
Sequence: 1 caggtatgttaaatgtgaga.....tcaatttaccacacaca 123

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PCITUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16	13.0	1278	4 US-09-050-159-123	Sequence 123, App
C 2	15	12.2	96	3 US-09-069-886-12	Sequence 12, Appl
C 3	15	12.2	101	3 US-09-069-886-27	Sequence 27, Appl
C 4	15	12.2	1099	2 US-08-871-033-2	Sequence 2, Appl
C 5	15	12.2	1099	2 US-09-212-167-2	Sequence 2, Appl
C 6	15	12.2	3171	5 PCR-US95-09261-1	Sequence 1, Appl
C 7	14	11.4	528	3 US-09-157-177-111	Sequence 111, App
C 8	14	11.4	564	4 US-09-328-111-223	Sequence 223, App
C 9	14	11.4	574	4 US-09-328-111-223	Sequence 297, App
C 10	14	11.4	750	4 US-08-845-256-2	Sequence 2, Appl
C 11	14	11.4	877	2 US-09-208-210-2	Sequence 2, Appl
C 12	14	11.4	877	2 US-08-989-289-2	Sequence 2, Appl
C 13	14	11.4	1021	3 US-08-714-918-70	Sequence 70, Appl
C 14	14	11.4	1021	4 US-09-265-315-70	Sequence 70, Appl
C 15	14	11.4	1021	4 US-09-265-315-70	Sequence 70, Appl
C 16	14	11.4	1021	4 US-09-265-315-70	Sequence 70, Appl
C 17	14	11.4	1620	2 US-08-874-186-37	Sequence 37, Appl
C 18	14	11.4	2165	1 US-08-222-289-1	Sequence 1, Appl
C 19	14	11.4	2433	1 US-08-248-021A-1	Sequence 1, Appl
C 20	14	11.4	2466	6 5194375-5	Patent No. 5194375
C 21	14	11.4	4259	2 US-08-816-155B-2	Sequence 2, Appl
C 22	14	11.4	4259	3 US-09-079-587-2	Sequence 2, Appl
C 23	14	11.4	4739	3 US-08-685-871-1	Sequence 1, Appl
C 24	14	11.4	5455	1 US-08-342-930-1	Sequence 3, Appl
C 25	14	11.4	8561	3 US-09-112-450-3	Sequence 1, Appl
C 26	14	11.4	10281	2 US-08-816-155B-1	Sequence 1, Appl
C 27	14	11.4	10281	3 US-09-079-587-1	Sequence 1, Appl

C 28	14	11.4	43795	3 US-08-742-185-101	Sequence 101, App
C 29	13	10.6	30	4 US-09-503-391-16	Sequence 16, Appl
C 30	13	10.6	33	4 US-08-427-569-14	Sequence 14, Appl
C 31	13	10.6	36	1 US-08-311-760A-315	Sequence 315, App
C 32	13	10.6	36	1 US-08-311-760A-315	Sequence 315, App
C 33	13	10.6	36	2 US-08-774-310-316	Sequence 316, App
C 34	13	10.6	36	2 US-08-774-310-316	Sequence 316, App
C 35	13	10.6	39	1 US-08-207-226A-3	Sequence 3, Appl
C 36	13	10.6	40	2 US-08-425-684-133	Sequence 133, App
C 37	13	10.6	40	2 US-08-675-502-133	Sequence 133, App
C 38	13	10.6	42	3 US-09-129-740-10	Sequence 10, Appl
C 39	13	10.6	50	2 US-08-821-559A-17	Sequence 17, Appl
C 40	13	10.6	70	5 PCR-US95-14418-50	Sequence 50, Appl
C 41	13	10.6	70	5 PCR-US95-15327-50	Sequence 50, Appl
C 42	13	10.6	71	4 US-08-569-147-69	Sequence 69, Appl
C 43	13	10.6	95	3 US-09-235-246-21	Sequence 21, Appl
C 44	13	10.6	96	1 US-08-253-877C-12	Sequence 12, Appl
C 45	13	10.6	96	2 US-08-452-164A-12	Sequence 12, Appl

## ALIGNMENTS

```
RESULT 1
US-09-050-159-123/C
; Sequence 123, Application US/09050159A
; Patent No. 6197505
; GENERAL INFORMATION:
; APPLICANT: NO. 6197505berg, Leif T
; APPLICANT: Andersson, Maria K
; APPLICANT: Linstrom, Per H
; TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND
; FILE REFERENCE: 1246/10042
; CURRENT APPLICATION NUMBER: US/09/050,159A
; EARLIER FILING DATE: 1998-03-27
; EARLIER APPLICATION NUMBER: 60/042,930
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 123
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Angiotensinogen, 5' region and exon 1
US-09-050-159-123

Query Match 13.0%; Score 16; DB 4; Length 1278;
Best local similarity 100.0%; Pred. No. 5.3;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 tggtaagttcacatg 86
Db 522 TGCTAGTTCACATG 507

RESULT 2
US-09-069-886-12
; Sequence 12, Application US/09069886
; Patent No. 6132724
; GENERAL INFORMATION:
; APPLICANT: Blum, Kenneth
; APPLICANT: Comings, David E.
; APPLICANT: Ivy, John L.
; TITLE OF INVENTION: ALLELIC POLYGENE DIAGNOSIS OF REWARD
; TITLE OF INVENTION: DEFICIENCY SYNDROME AND TREATMENT
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
```



STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/069,886  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hodgins, Daniel S.  
REGISTRATION NUMBER: 31,026  
REFERENCE/DOCKET NUMBER: BLUM:002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512)418-3000  
TELEFAX: (512)474-7577  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-069-886-12

Query Match 12.2%; Score 15; DB 3; Length 96;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtg 17  
|||||  
DB 15 GGTATTGTAATGTC 29

RESULT 3  
US-09-069-886-27  
Sequence 27, Application US/09069886  
Patent No. 6132724  
GENERAL INFORMATION:  
APPLICANT: Blum, Kenneth  
APPLICANT: Comings, David E.  
APPLICANT: Ivy, John L.  
TITLE OF INVENTION: ALLELIC POLYGENE DIAGNOSIS OF REWARD  
TITLE OF INVENTION: DEFICIENCY SYNDROME AND TREATMENT  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/069,886  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hodgins, Daniel S.  
REGISTRATION NUMBER: 31,026  
REFERENCE/DOCKET NUMBER: BLUM:002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512)418-3000  
TELEFAX: (512)474-7577  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:

LENGTH: 101 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-069-886-27

Query Match 12.2%; Score 15; DB 3; Length 101;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtg 17  
|||||  
DB 15 GGTATTGTAATGTC 29

RESULT 4  
US-08-871-033-2/c  
Sequence 2, Application US/08871033  
Patent No. 5869259  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NEW CARBOXYPEPTIDASE INHIBITOR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,033  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0318 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1099 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT09  
CLONE: 1649584  
US-08-871-033-2

Query Match 12.2%; Score 15; DB 2; Length 1099;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 taagtgataactat 101  
|||||  
DB 307 TAAGTGATACTTAT 293



RESULT 5  
US-09-212-167-2/c  
; Sequence 2, Application US/09212167  
; Patent No. 5998373  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; TITLE OF INVENTION: NEW CARBOXYPEPTIDASE INHIBITOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/212,167  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/871,033  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0318 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1099 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PROSTUT09  
; CLONE: 1649584  
; US-09-212-167-2

Query Match 12.2%; Score 15; DB 2; Length 1099;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 taagtgatacttat 101  
|||||

DB 307 TAAGTGATACCTTAT 293

RESULT 6  
PCT-US95-09261-1/c  
; Sequence 1, Application PC/TUS9509261  
; GENERAL INFORMATION:  
; APPLICANT:  
; APPLICANT: NAME: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM  
; STREET: 201 West 7th Street  
; CITY: Austin  
; APPLICANT: STATE: Texas  
; APPLICANT: COUNTRY: United States of America  
; APPLICANT: POSTAL CODE: 78701  
; APPLICANT: TELEPHONE NO.: (512)499-4462  
; APPLICANT: TELEFAX: (512)499-4523  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE EXPRESSION OF  
; TITLE OF INVENTION: A BONE AND PROSTATE DERIVED GROWTH FACTOR  
; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/09261  
; FILING DATE: CONCURRENTLY HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/283,701  
; FILING DATE: 01-AUG-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SERTICH, GARY J.  
; REGISTRATION NUMBER: 34,430  
; REFERENCE/DOCKET NUMBER: UTFC422P--  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; TELEX: 79-0824  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3171 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "DNA"  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 694..2310  
; PCT-US95-09261-1

Query Match 12.2%; Score 15; DB 5; Length 3171;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 aattattccacaca 120  
|||||

DB 3143 AATTATTCCACACA 3129

RESULT 7  
US-09-157-177-111/c  
; Sequence 111, Application US/09157177  
; Patent No. 6090558  
; GENERAL INFORMATION:  
; APPLICANT: Butler, John M.  
; APPLICANT: Li, Jia  
; APPLICANT: Monforte, Joseph A.  
; APPLICANT: Becker, Christopher H.  
; TITLE OF INVENTION: DNA TYPING BY MASS SPECTROMETRY WITH POLYMORPHIC DNA  
; TITLE OF INVENTION: REPEAT MARKERS  
; FILE REFERENCE: GSTR:017/GSTR017P  
; CURRENT APPLICATION NUMBER: US/09/157,177  
; CURRENT FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: 60/059,415  
; EARLIER FILING DATE: 1997-09-19  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 111  
; LENGTH: 528  
; TYPE: DNA  
; ORGANISM: Homo sapiens



US-09-157-177-111

Query Match 11.4%; Score 14; DB 3; Length 528;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 atgtgagatgcctt 26  
|||||  
Db 46 ATGTGAGATGCCTT 33

RESULT 8

US-09-328-111-223  
; Sequence 223, Application US/09328111  
; Patent No. 6262333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Dertli, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; CURRENT FILING DATE: 1999-06-08  
; EARLIER APPLICATION NUMBER: US 60/088,801  
; EARLIER FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 223  
; LENGTH: 564  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(564)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-223

Query Match 11.4%; Score 14; DB 4; Length 564;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 tcttcattatttc 114  
|||||  
Db 117 tcttcattatttc 130

RESULT 9

US-09-328-111-297  
; Sequence 297, Application US/09328111  
; Patent No. 6262333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Dertli, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; CURRENT FILING DATE: 1999-06-08  
; EARLIER APPLICATION NUMBER: US 60/088,801  
; EARLIER FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 297  
; LENGTH: 574  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(574)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-297

Query Match 11.4%; Score 14; DB 4; Length 574;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 tcttcattatttc 114  
|||||  
Db 144 tcttcattatttc 157

RESULT 10

US-08-845-256-2  
; Sequence 2, Application US/08845256  
; Patent No. 5874244  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: NEW HUMAN NMDH DEHYDROGENASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/845,256  
; FILING DATE: Filed Herewith  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0277 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 750 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BEPIWOT01



```
; CLONE: 2054787
US-08-845-256-2

Query Match          11.4%; Score 14; DB 2; Length 750;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ttgtaaatgtgaga 20
   |||||||
Db 641 TTGTAATGTGAGA 654

RESULT 11
US-09-208-210-2
; Sequence 2, Application US/09208210
; Patent No. 6277572
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NEW HUMAN NADH DEHYDROGENASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/208,210
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,256
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0277 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BEP1NOT01
; CLONE: 2054787
; US-09-208-210-2

Query Match          11.4%; Score 14; DB 4; Length 750;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ttgtaaatgtgaga 20
   |||||||
Db 641 TTGTAATGTGAGA 654

RESULT 12
US-08-989-289-2
; Sequence 2, Application US/08989289
; Patent No. 5968747
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: UBIQUITIN-LIKE CONJUGATING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,289
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0439 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 877 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: ADREUT05
; CLONE: 2501808
; US-08-989-289-2

Query Match          11.4%; Score 14; DB 2; Length 877;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 caacctgatgacc 41
   |||||||
Db 449 CAACCTGATGACC 462

RESULT 13
US-08-714-918-70/C
; Sequence 70, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ying
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmidt, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
```



COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/714,918  
FILING DATE: September 13, 1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 222/005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ. ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1021 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-714-918-70

Query Match 11.4%; Score 14; DB 3; Length 1021;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaattgt 16  
|||||  
Db 523 GGTATTGTAATGT 510

RESULT 14  
US-09-265-315-70/c  
Sequence 70, Application US/09265315  
Patent No. 6187541  
GENERAL INFORMATION:  
APPLICANT: Benton, Bret  
APPLICANT: Lee, Ying J.  
APPLICANT: Malouin, Francois  
APPLICANT: Martin, Patrick K.  
APPLICANT: Schmid, Molly B.  
APPLICANT: Sun, Dongxu  
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/265,315  
FILING DATE: March 9, 1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/714,918  
FILING DATE: September 13, 1996  
APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 240/247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ. ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1021 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-265-315-70

Query Match 11.4%; Score 14; DB 4; Length 1021;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaattgt 16  
|||||  
Db 523 GGTATTGTAATGT 510

RESULT 15  
US-09-265-315-70/c  
Sequence 70, Application US/09265315  
Patent No. 6187541  
GENERAL INFORMATION:  
APPLICANT: Benton, Bret  
APPLICANT: Lee, Ying J.  
APPLICANT: Malouin, Francois  
APPLICANT: Martin, Patrick K.  
APPLICANT: Schmid, Molly B.  
APPLICANT: Sun, Dongxu  
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/265,315  
FILING DATE: March 9, 1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/714,918  
FILING DATE: September 13, 1996  
APPLICATION NUMBER: 60/009,102



FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 240/247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1021 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-265-315-70

Query Match 11.4%; Score 14; DB 4; Length 1021;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gqatgttaatgt 16  
|||||  
DB 523 GGTATTGTAATGT 510

RESULT 16  
US-09-266-417-70/c  
Sequence 70, Application US/09266417  
Patent No. 6228588  
GENERAL INFORMATION:  
APPLICANT: Benton, Bret  
APPLICANT: Lee, Ying J.  
APPLICANT: Malouin, Francois  
APPLICANT: Martin, Patrick K.  
APPLICANT: Schmid, Molly B.  
APPLICANT: Sun, Dongxu  
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
TITLE OF INVENTION: TARGET GENES  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/266,417  
FILING DATE: March 9, 1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/714,918  
FILING DATE: September 13, 1996  
APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 240/248

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1021 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-266-417-70

Query Match 11.4%; Score 14; DB 4; Length 1021;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gqatgttaatgt 16  
|||||  
DB 523 GGTATTGTAATGT 510

RESULT 17  
US-08-874-186-37/c  
Sequence 37, Application US/08874186  
Patent No. 5989885  
GENERAL INFORMATION:  
APPLICANT: Teng, David H-F.  
APPLICANT: Tavligian, Sean V.  
APPLICANT: Perry III, William L.  
APPLICANT: Skolnick, Mark H.  
TITLE OF INVENTION: SPECIFIC MUTATIONS OF MAP KINASE KINASE  
TITLE OF INVENTION: 4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR  
NUMBER OF SEQUENCES: 96  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, N.W., Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/874,186  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/782,482  
FILING DATE: 10-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Stephen A.  
REGISTRATION NUMBER: 38,609  
REFERENCE/DOCKET NUMBER: 24884-121392-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4848  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1620 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 1..1325



FEATURE:  
NAME/KEY: exon  
LOCATION: 1326..1428  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 1429..1620  
US-08-874-186-37

Query Match 11.4%; Score 14; DB 2; Length 1620;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ttttaattgagat 21  
|||||  
DB 269 TGTGAGATGAGAT 256

RESULT 18  
US-08-222-289-1/c  
Sequence 1, Application US/08222289  
Patent No. 5559010  
GENERAL INFORMATION:  
APPLICANT: Kilonsky, Daniel J  
APPLICANT: Destruelle, Monika  
APPLICANT: Holzer, Helmut  
TITLE OF INVENTION: NUTRIENT REGULATED GENE EXPRESSION  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHNACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/222,289  
FILING DATE: 04-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard Aron.  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59767/RMO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2165 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 370..1434  
US-08-222-289-1

Query Match 11.4%; Score 14; DB 1; Length 2165;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 14 tttgagatgcctta 27  
|||||  
DB 475 TGTGAGATGCTTA 462

RESULT 19  
US-08-248-021A-1/c  
Sequence 1, Application US/08248021A  
Patent No. 5648240  
GENERAL INFORMATION:  
APPLICANT: Hook, Magnus  
APPLICANT: Jonsson, Klas  
APPLICANT: Patti, Joseph M.  
APPLICANT: Gurusiddappa, Sivashankarappa  
TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/248,021A  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: TANK:155  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2433 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-248-021A-1

Query Match 11.4%; Score 14; DB 1; Length 2433;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 tttgagatgcctta 27  
|||||  
DB 2169 TGTGAGATGCTTA 2156

RESULT 20  
5194375-5  
Patent No. 5194375  
APPLICANT: PARK, LINDA S.;GOODWIN, RAYMOND G.  
TITLE OF INVENTION: DNA ENCODING INTERLEUKIN-7 RECEPTORS AND  
METHODS OF USE  
NUMBER OF SEQUENCES: 9  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/493,588  
FILING DATE: 21-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 421,201  
FILING DATE: 13-OCT-1989  
APPLICATION NUMBER: 366,910  
FILING DATE: 15-JUN-1989  
SEQ ID NO:5:  
LENGTH: 2466  
5194375-5



Query Match 11.4%; Score 14; DB 6; Length 2466;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 ttattccacacaa 121  
|||||  
Db 593 ttattccacacaa 606

## RESULT 21

US-08-816-155B-2/c  
; Sequence 2, Application US/08816155B  
; Patent No. 5990091  
; GENERAL INFORMATION:  
; APPLICANT: TARTAGLIA, JAMES  
; APPLICANT: COX, WILLIAM I.  
; APPLICANT: GETTIG, RUSSELL R.  
; APPLICANT: MARTINEZ, HECTOR  
; APPLICANT: PAOLETTI, ENZO  
; APPLICANT: PINCUS, STEVEN E.  
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
; TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
; STREET: 745 FIFTH AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10151  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,155B  
; FILING DATE: 12-MAR-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOMALSKI, THOMAS J.  
; REGISTRATION NUMBER: 32,147  
; REFERENCE/DOCKET NUMBER: 454310-2990  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-588-0800  
; TELEFAX: 212-588-0500  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4259 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-816-155B-2

## Query Match

Best Local Similarity 11.4%; Score 14; DB 2; Length 4259;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 107 attattccacaca 120  
|||||  
Db 1759 ATTATTCCACACA 1746

## RESULT 22

US-09-079-587-2/c  
; Sequence 2, Application US/09079587  
; Patent No. 6130066  
; GENERAL INFORMATION:  
; APPLICANT: TARTAGLIA, JAMES  
; APPLICANT: COX, WILLIAM I.  
; APPLICANT: GETTIG, RUSSELL R.

APPLICANT: MARTINEZ, HECTOR  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: PINCUS, STEVEN E.  
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
STREET: 745 FIFTH AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10151

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,587  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/816,155  
FILING DATE: 12-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: KOMALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2990  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4259 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-079-587-2

Query Match 11.4%; Score 14; DB 3; Length 4259;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 107 attattccacaca 120  
|||||  
Db 1759 ATTATTCCACACA 1746

## RESULT 23

US-08-685-871-1  
; Sequence 1, Application US/08685871  
; Patent No. 6013499  
; GENERAL INFORMATION:  
; APPLICANT: NARUMIYA, Shuh  
; APPLICANT: IWAMATSU, Akihiro  
; TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160  
; NUMBER OF SEQUENCES: 68  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:



```

APPLICATION NUMBER: US/08/685,871
FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184102
FILING DATE: 25-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-262553
FILING DATE: 14-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16887/845
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4739 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 448..4509
US-08-685-871-1

```

```

Query Match          11.4% Score 14; DB 3; Length 4739;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 93 gatactatctca 106
    |||||||||||
Db 1658 GATACTACTCTCA 1671

```

```

RESULT 24
US-08-342-930-1
Sequence 1, Application US/08342930
Patent No. 5821084
GENERAL INFORMATION:
APPLICANT: OLIMSTED, ELIZABETH A.
APPLICANT: MAURO, LAURA J.
APPLICANT: DAVIS, ALAN R.
APPLICANT: DIXON, JACK E.
TITLE OF INVENTION: OSTEOBLAST-TESTICULAR PROTEIN TYROSINE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,930
FILING DATE: 21-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KONSRI, ANTOINETTE F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: 20344-20975.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792

```

```

TELEX: 706141
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5455 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 205..5337
US-08-342-930-1

```

```

Query Match          11.4% Score 14; DB 1; Length 5455;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 19 gatgccttaacc 32
    |||||||||||
Db 3558 GATGCTTACACC 3571

```

```

RESULT 25
US-09-112-450-3/c
Sequence 3, Application US/09112450
Patent No. 6120999
GENERAL INFORMATION:
APPLICANT: Abad, Antonio Jose C.
APPLICANT: Choi, Gil
APPLICANT: Caldeone, Richard A.
TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
FILE REFERENCE: PF393
CURRENT APPLICATION NUMBER: US/09/112,450
CURRENT FILING DATE: 1998-07-09
EARLIER APPLICATION NUMBER: 60/052,273
EARLIER FILING DATE: 1997-07-10
EARLIER APPLICATION NUMBER: 60/074,308
EARLIER FILING DATE: 1998-02-11
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 8561
TYPE: DNA
ORGANISM: Homo sapiens
US-09-112-450-3

```

```

Query Match          11.4% Score 14; DB 3; Length 8561;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 aggtatgtaaatg 15
    |||||||||||
Db 1667 AGGTATGTAAATG 1654

```

```

RESULT 26
US-08-816-155B-1/c
Sequence 1, Application US/08816155B
Patent No. 5990091
GENERAL INFORMATION:
APPLICANT: TARTAGLIA, JAMES
APPLICANT: COX, WILLIAM I.
APPLICANT: GETTIG, RUSSELL R.
APPLICANT: MARTINEZ, HECTOR
APPLICANT: PROLETTI, ENZO
APPLICANT: PINCUS, STEVEN E.
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
METHODS OF MAKING AND USING THEREOF
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESS: FROMMER LAWRENCE & HAUG LLP
STREET: 745 FIFTH AVENUE

```



CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10151  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,155B  
FILING DATE: 12-MAR-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: KOMALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2990  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10281 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-816-155B-1

Query Match 11.4%; Score 14; DB 2; Length 10281;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 attattccacaca 120  
|||||  
DB 1788 ATTATTCCACACA 1775

RESULT 27  
US-09-079-587-1/c  
Sequence 1, Application US/09079587  
Patent No. 6130066  
GENERAL INFORMATION:  
APPLICANT: TARTAGLIA, JAMES  
APPLICANT: COX, WILLIAM I.  
APPLICANT: GETTIG, RUSSELL R.  
APPLICANT: MARTINEZ, HECTOR  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: PINCUS, STEVEN E.  
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
STREET: 745 FIFTH AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10151  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,587  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/816,155  
FILING DATE: 12-MAR-1997  
ATTORNEY/AGENT INFORMATION:

NAME: KOMALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2990  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10281 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-079-587-1

Query Match 11.4%; Score 14; DB 3; Length 10281;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 attattccacaca 120  
|||||  
DB 1788 ATTATTCCACACA 1775

RESULT 28  
US-08-742-185-101  
Sequence 101, Application US/08742185  
Patent No. 6020476  
GENERAL INFORMATION:  
APPLICANT: Page, David C.  
APPLICANT: Reijo, Renee  
APPLICANT: Saxena, Richa  
APPLICANT: Hawkins, Trevor  
APPLICANT: Reeve, Mary Pat  
TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/742,185  
FILING DATE: 30-OCT-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/690,734  
FILING DATE: 31-JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/310,429  
FILING DATE: 22-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WH194-07A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 101:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43795 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)



US-08-742-185-101

Query Match 11.4%; Score 14; DB 3; Length 43795;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 attcattattt 113  
|||||  
Db 32551 ATCTCATTTATT 32564

RESULT 29

US-09-503-391-16/c  
Sequence 16, Application US/09503391  
Patent No. 6300091  
GENERAL INFORMATION:  
APPLICANT: Patton, David A.  
APPLICANT: Ashby, Carl S.  
APPLICANT: Thomas, Carla R.  
APPLICANT: McEliver, John A.  
APPLICANT: Budziszewski, Gregory J.  
APPLICANT: Levin, Joshua Z.  
TITLE OF INVENTION: Herbicide Target Genes and Methods  
FILE REFERENCE: PB/5-30852A  
CURRENT FILING DATE: 2000-02-14  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 16  
LENGTH: 30  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-09-503-391-16

Query Match 10.6%; Score 13; DB 4; Length 30;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 attccacacaca 123  
|||||  
Db 16 ATTCACACACACA 4

RESULT 30

US-08-427-569-14/c  
Sequence 14, Application US/08427569  
Patent No. 6235465  
GENERAL INFORMATION:  
APPLICANT: Kolberg, Janice A.  
APPLICANT: Urdea, Michael S.  
TITLE OF INVENTION: HTLV-1 PROBES FOR USE IN SOLUTION  
NUMBER OF INVENTIONS: PHASE SANDWICH HYBRIDIZATION ASSAYS  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/427,569  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/130,150  
FILING DATE:  
APPLICATION NUMBER: 07/813,585  
FILING DATE: 23-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Thomas E. Clotf1  
REGISTRATION NUMBER: 21,013  
REFERENCE/DOCKET NUMBER: 22300-20238.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-813-5600  
TELEFAX: 415-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-427-569-14

Query Match 10.6%; Score 13; DB 4; Length 33;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 gatgccttaccac 31  
|||||  
Db 18 GATGCTTACAC 6

RESULT 31

US-08-311-760A-315  
Sequence 315, Application US/08311760A  
Patent No. 5599706  
GENERAL INFORMATION:  
APPLICANT: Stinchcomb, Dan T.  
APPLICANT: McSwiggen, James  
APPLICANT: Newton, Roger S.  
APPLICANT: Ramharack, Randy  
TITLE OF INVENTION: RIBOZYME TREATMENT OF DISEASES  
TITLE OF INVENTION: OR CONDITIONS RELATED TO LEVELS OF  
TITLE OF INVENTION: PLASMA LIPOPROTEIN (a) [LP(a)] BY  
TITLE OF INVENTION: INHIBITING APOLOPROTEIN  
NUMBER OF SEQUENCES: 392  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,760A  
FILING DATE: September 23, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 208/155  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600



TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 315:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-311-760A-315

Query Match 10.6%; Score 13; DB 1; Length 36;  
Best Local Similarity 84.6%; Pred. No. 2.3e+02;  
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 27 acaaccctgatga 39  
|||||||:|:|:|  
Db 2 ACAACCCUGAUGA 14

## RESULT 32

US-08-311-760A-316  
Sequence 316, Application US/08311760A  
Patent No. 5599706  
GENERAL INFORMATION:  
APPLICANT: Stinchcomb, Dan T.  
APPLICANT: McSwigen, James  
APPLICANT: Newton, Roger S.  
TITLE OF INVENTION: RIBOZYME TREATMENT OF DISEASES  
TITLE OF INVENTION: OR CONDITIONS RELATED TO LEVELS OF  
TITLE OF INVENTION: PLASMA LIPOPROTEIN (a) [LP(a)] BY  
TITLE OF INVENTION: INHIBITING APOLIPOPROTEIN  
NUMBER OF SEQUENCES: 392  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,760A  
FILING DATE: September 23, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 208/155  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 316:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-311-760A-316

Query Match 10.6%; Score 13; DB 1; Length 36;

Best Local Similarity 84.6%; Pred. No. 2.3e+02;  
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 27 acaaccctgatga 39  
|||||||:|:|:|  
Db 2 ACAACCCUGAUGA 14

## RESULT 33

US-08-774-310-315  
Sequence 315, Application US/08774310  
Patent No. 5877022  
GENERAL INFORMATION:  
APPLICANT: Stinchcomb, Daniel T.  
APPLICANT: McSwigen, James  
APPLICANT: Newton, Roger S.  
TITLE OF INVENTION: RIBOZYME TREATMENT OF DISEASES  
TITLE OF INVENTION: OR CONDITIONS RELATED TO LEVELS OF  
TITLE OF INVENTION: PLASMA LIPOPROTEIN (a) [LP(a)] BY  
TITLE OF INVENTION: INHIBITING APOLIPOPROTEIN  
NUMBER OF SEQUENCES: 392  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/774,310  
FILING DATE: December 23, 1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/311,760  
FILING DATE: September 23, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 223/229  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 315:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-774-310-315

Query Match 10.6%; Score 13; DB 2; Length 36;  
Best Local Similarity 84.6%; Pred. No. 2.3e+02;  
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 27 acaaccctgatga 39  
|||||||:|:|:|  
Db 2 ACAACCCUGAUGA 14

## RESULT 34

US-08-774-310-316  
Sequence 316, Application US/08774310  
Patent No. 5877022



GENERAL INFORMATION:  
APPLICANT: Scinchcomb, Daniel T.  
APPLICANT: McSwigen, James  
APPLICANT: Newton, Roger S.  
APPLICANT: Ramharack, Randy  
TITLE OF INVENTION: RIBOZYME TREATMENT OF DISEASES  
TITLE OF INVENTION: OR CONDITIONS RELATED TO LEVELS OF  
TITLE OF INVENTION: PLASMA LIPOPROTEIN (a). [LP(a)] BY  
TITLE OF INVENTION: INHIBITING APOLIPOPROTEIN  
TITLE OF INVENTION:  
NUMBER OF SEQUENCES: 392  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: Storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/774,310  
FILING DATE: December 23, 1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/311,760  
FILING DATE: September 23, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Wardburg, Richard  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 223/229  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 316:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-774-310-316

Query Match 10.6%; Score 13; DB 2; Length 36;  
Best Local Similarity 84.6%; Pred. No. 2.3e+02;  
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 27 acaaccctgatga 39  
|||||:|:|  
Db 2 ACAACCCUGAUGA 14

RESULT 35  
US-08-207-226A-3/C  
Sequence 3, Application US/08207226A  
Patent No. 5580970  
GENERAL INFORMATION:  
APPLICANT: Hendricks, David  
APPLICANT: Rigby, Susan  
APPLICANT: Parodos, Kyriaki  
TITLE OF INVENTION: DETECTION OF HPV TRANSCRIPTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corporation  
STREET: 55 Shuman Boulevard, Suite 600  
CITY: Naperville  
STATE: Illinois  
COUNTRY: US

ZIP: 60563  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/207,226A  
FILING DATE: 04-MAR-1994  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/622,742  
FILING DATE: 05-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5580970val B.  
REFERENCE/DOCKET NUMBER: GTR-8903AF  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-207-226A-3

Query Match 10.6%; Score 13; DB 1; Length 39;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ttgtaattgtgag 19  
|||||  
Db 30 TTGTAATGTGAG 18

RESULT 36  
US-08-425-684-133/C  
Sequence 133, Application US/08425684  
Patent No. 5834252  
GENERAL INFORMATION:  
APPLICANT: STEMMER PH.D., WILLEM P.C.  
TITLE OF INVENTION: END-COMPLEMENTARY POLYMERASE REACTION  
NUMBER OF SEQUENCES: 136  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW  
STREET: ONE MARKET PLAZA, STEUART STREET TOWER, 20TH FLOOR  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,684  
FILING DATE: 18-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: DUNN ESQ., TRACY J.  
REGISTRATION NUMBER: 34,587  
REFERENCE/DOCKET NUMBER: 165284-015400US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 133:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
US-08-425-684-133



Query Match 10.6%; Score 13; DB 2; Length 40;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123  
Db 21 ATTCCACACACACA 9

## RESULT 37

US-08-675-502-133/c  
; Sequence 133, Application US/08675502  
; Patent No. 5928905  
; GENERAL INFORMATION:  
; APPLICANT: STEMMER PH.D., WILLEM P.C.  
; APPLICANT: LIPSHUTZ, ROBERT J.  
; TITLE OF INVENTION: END-COMPLEMENTARY POLYMERASE REACTION  
; NUMBER OF SEQUENCES: 136  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW  
; STREET: TWO EMBARCADERO CENTER, 8TH FLOOR  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/675,502  
; FILING DATE: 03-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/425,684  
; FILING DATE: 18-APR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/05480  
; FILING DATE: 18-APR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY, ESQ., MATTHEW B.  
; REGISTRATION NUMBER: 39,787  
; REFERENCE/DOCKET NUMBER: 16528J-015410US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 326-2400  
; TELEFAX: (415) 326-2422  
; INFORMATION FOR SEQ ID NO: 133:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; US-08-675-502-133

Query Match 10.6%; Score 13; DB 2; Length 40;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123  
Db 21 ATTCCACACACACA 9

## RESULT 38

US-09-129-740-10/c  
; Sequence 10, Application US/09129740A  
; Patent No. 6067246

; GENERAL INFORMATION:  
; APPLICANT: HELLER, MICHAEL J.  
; APPLICANT: TU, EUGENE  
; TITLE OF INVENTION: DNA OPTICAL STORAGE  
; FILE REFERENCE: DAVID B. MURPHY - Heller 236/119  
; CURRENT APPLICATION NUMBER: US/09/129,740A  
; CURRENT FILING DATE: 1998-08-05  
; EARLIER APPLICATION NUMBER: 08/906,569  
; EARLIER FILING DATE: 1997-08-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 42  
; TYPE: DNA  
; ORGANISM: synthetic construct  
; US-09-129-740-10

Query Match 10.6%; Score 13; DB 3; Length 42;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123  
Db 30 ATTCCACACACACA 18

## RESULT 39

US-08-821-559A-17/c  
; Sequence 17, Application US/08821559A  
; Patent No. 5846774  
; GENERAL INFORMATION:  
; APPLICANT: XIA, YUANNAN  
; TITLE OF INVENTION: CHLORELLA VIRUS PROMOTERS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
; STREET: 3100 No. 5846774west Center, 90 South Seventh st  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/821,559A  
; FILING DATE: 21-MAR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kettelberger, Denise M  
; REGISTRATION NUMBER: 33,924  
; REFERENCE/DOCKET NUMBER: 8648.63-US-01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612/371-5268  
; TELEFAX: 612/332-9081  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-821-559A-17

Query Match 10.6%; Score 13; DB 2; Length 50;



Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123  
|||||  
Db 49 ATTCCACACACA 37

## RESULT 40

PCT-US95-14418-50  
Sequence 50, Application PC/TUS9514418  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DNA Encoding a Thermostable DNA Polymerase Enzyme  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/14418  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gass, David A.  
REGISTRATION NUMBER: 38,153  
REFERENCE/DOCKET NUMBER: 28003/32330  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 70 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US95-14418-50

Query Match 10.6%; Score 13; DB 5; Length 70;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123  
|||||  
Db 57 ATTCCACACACA 69

## RESULT 41

PCT-US95-15327-50  
Sequence 50, Application PC/TUS9515327  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Biologically Active Fragments of  
TITLE OF INVENTION: Therms Flavus DNA Polymerase  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/15327  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gass, David A.  
REGISTRATION NUMBER: 38,153  
REFERENCE/DOCKET NUMBER: 28003/31716  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 70 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US95-15327-50

Query Match 10.6%; Score 13; DB 5; Length 70;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123  
|||||  
Db 57 ATTCCACACACA 69

RESULT 42  
US-08-569-147-69  
Sequence 69, Application US/08569147  
Patent No. 6180377  
GENERAL INFORMATION:

APPLICANT:  
TITLE OF INVENTION: HUMANISED ANTIBODIES  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
ADDRESSEE: No. 6180377ris, LLP  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: U.S.A.  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/569,147  
FILING DATE: 25-March-1996  
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:  
NAME: Trujillo, Doreen Yanko  
REGISTRATION NUMBER: 35,719  
REFERENCE/DOCKET NUMBER: CARP-0047  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 71 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear



MOLECULE TYPE: CDNA  
US-08-569-147-69

Query Match 10.6%; Score 13; DB 4; Length 71;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 agatgccttaca 30  
DB 28 AGATGCCTTACAA 40

RESULT 43

US-09-235-246-21/C  
Sequence 21, Application US/09235246A  
Patent No. 6048719

GENERAL INFORMATION:

APPLICANT: Kong, Huimin  
APPLICANT: Higgins, Lauren S.  
APPLICANT: Dalton, Michael A.  
TITLE OF INVENTION: Method For Cloning And Producing The DraIII Restriction  
FILE REFERENCE: DraIII  
CURRENT APPLICATION NUMBER: US/09/235,246A  
CURRENT FILING DATE: 1999-01-22  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 21  
LENGTH: 95  
TYPE: DNA  
ORGANISM: Deinococcus radiophilus  
US-09-235-246-21

Query Match 10.6%; Score 13; DB 3; Length 95;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123  
DB 57 ATTCCACACACA 45

RESULT 44

US-08-253-877C-12/C  
Sequence 12, Application US/08253877C  
Patent No. 5773001

GENERAL INFORMATION:

APPLICANT: Hamann, Phillip R.  
APPLICANT: Hinman, Lois  
APPLICANT: Hollander, Irwin  
APPLICANT: Holcomb, Ryan  
APPLICANT: Hallett, William  
APPLICANT: Tsou, Hwei-Ru  
APPLICANT: Weiss, Martin J.  
TITLE OF INVENTION: Conjugates of Methyltrithio Antitumor  
TITLE OF INVENTION: Agents and Intermediates for Their Synthesis  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07470-8426

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/253,877C

FILING DATE: 03-JUN-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 32,368

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3246  
TELEFAX: 201-831-3305

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:

LENGTH: 96 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-253-877C-12

Query Match 10.6%; Score 13; DB 1; Length 96;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 agatgccttaca 30  
DB 78 AGATGCCTTACAA 66

RESULT 45

US-08-452-164A-12/C  
Sequence 12, Application US/08452164A  
Patent No. 5877296

GENERAL INFORMATION:

APPLICANT: Hamann, Phillip R.  
APPLICANT: Hinman, Lois  
APPLICANT: Hollander, Irwin  
APPLICANT: Holcomb, Ryan  
APPLICANT: Hallett, William  
APPLICANT: Tsou, Hwei-Ru  
APPLICANT: Weiss, Martin J.  
TITLE OF INVENTION: Conjugates of Methyltrithio Antitumor  
TITLE OF INVENTION: Agents and Intermediates for Their Synthesis  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Home Products Corporation  
STREET: One Campus Drive  
CITY: Parsippany  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07054

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,164A  
FILING DATE: 26-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 32,368-04  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-683-2158  
TELEFAX: 201-683-4117

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)



US-08-452-164A-12

Query Match 10.6%; Score 13; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 18 agatgccttaca 30  
|||||  
Db 78 AGATGCCTTACAA 66

Search completed: March 19, 2002, 11:04:27  
Job time: 3856 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:31.36 ; Search time 1345.18 Seconds  
(without alignments)  
1508.462 Million cell updates/sec

Title: US-09-684-016-48411  
Perfect score: 123  
Sequence: 1 caggtatgtgaatgtaga.....tcaattatccacacaca 123

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_ph:\*  
7: gb\_pl:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: em\_ba:\*  
15: em\_fun:\*  
16: em\_hum:\*  
17: em\_in:\*  
18: em\_om:\*  
19: em\_or:\*  
20: em\_ov:\*  
21: em\_pat:\*  
22: em\_ph:\*  
23: em\_pl:\*  
24: em\_ro:\*  
25: em\_sts:\*  
26: em\_sy:\*  
27: em\_un:\*  
28: em\_vl:\*  
29: em\_hlg\_hum:\*  
30: em\_hlg\_inv:\*  
31: em\_hlg\_rod:\*  
32: em\_hlg\_rod:\*  
33: em\_hlg\_rod:\*  
34: em\_hlg\_rod:\*  
35: em\_hlg\_rod:\*  
36: em\_hlg\_rod:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	54.4	44.2	89035	8	ATT22F8	AL050351 Arabidops
C 2	54.4	44.2	100469	8	ATF19H22	AL035679 Arabidops
C 3	54.4	44.2	198987	8	ATCHRIV90	AL161594 Arabidops
C 4	53.4	43.4	103270	8	ATF1N20	AL022140 Arabidops
C 5	53.4	43.4	196247	8	ATCHRIV56	AL161556 Arabidops
C 6	52.4	42.6	119807	8	AP003753	AP003753 Oryza sat
C 7	47.4	38.2	1062	8	AF277453	AF277453 Arabidops
C 8	45.4	36.9	1058	8	HVES43	X77575 H. vulgare (
C 9	41.8	34.0	367	8	AB011445	AB011445 Trilicium
C 10	34.4	28.0	44723	6	AX059527	AX059527 Sequence
C 11	34.4	28.0	80446	8	T27D20	AF076274 Arabidops
C 12	34.4	28.0	198022	8	ATCHRIV11	AL161773 Arabidops
C 13	33.8	27.5	154195	9	AL161773	AL161773 Human DNA
C 14	33.8	27.5	166266	2	AP001146	AP001146 Homo sapi
C 15	33.8	27.5	185404	2	AC025409	AC025409 Homo sapi
C 16	31.6	25.7	265815	2	AC027308	AC027308 Homo sapi
C 17	31.4	25.5	799	5	GGGP23	X60795 Chicken mRN
C 18	31.2	25.4	23366	1	AE002311	AE002311 Chlamydia
C 19	31	25.2	30494	3	CEK1006	Z70440 Caenorhabdl
C 20	30	24.4	83145	8	AC002339	AC002339 Arabidops
C 21	29.8	24.2	68575	2	AC027346	AC027346 Homo sapi
C 22	29.8	24.2	161818	2	AP003409	AP003409 Oryza sat
C 23	29.8	24.2	170020	8	AP003256	AP003256 Oryza sat
C 24	29.6	24.1	63904	2	AC017284	AC017284 Drosophila
C 25	29.6	24.1	80065	2	AC016271	AC016271 Homo sapi
C 26	29.6	24.1	94797	3	AC005333	AC005333 Drosophila
C 27	29.6	24.1	166441	3	AE003239	AE003239 Drosophila
C 28	29.6	24.1	259973	3	AE003611	AE003611 Drosophila
C 29	29.4	23.9	155172	2	AC021110	AC021110 Homo sapi
C 30	29.4	23.9	168656	9	AL359922	AL359922 Human DNA
C 31	29.4	23.9	174815	2	AP001350	AP001350 Homo sapi
C 32	29.4	23.9	178056	2	AC091589	AC091589 Homo sapi
C 33	29.4	23.9	195491	2	AL449423	AL449423 Homo sapi
C 34	29.2	23.7	120515	9	AL136230	AL136230 Human DNA
C 35	29.2	23.7	159073	2	AC013621	AC013621 Homo sapi
C 36	29.2	23.7	159150	2	AC022863	AC022863 Homo sapi
C 37	29.2	23.7	170022	2	AC021559	AC021559 Homo sapi
C 38	29.2	23.7	181127	2	AC068203	AC068203 Homo sapi
C 39	29.2	23.7	181542	2	AC023187	AC023187 Homo sapi
C 40	28.8	23.4	1086	8	VAPPCGENE	X91634 V. aplylla m
C 41	28.8	23.4	14039	2	AC018023	AC018023 Drosophila
C 42	28.8	23.4	17078	1	AE001152	AE001152 Borrelia
C 43	28.8	23.4	49759	3	AE002717	AE002717 Drosophila
C 44	28.8	23.4	74534	3	AC004735	AC004735 Drosophila
C 45	28.8	23.4	85195	9	AL136093	AL136093 Human DNA

## ALIGNMENTS

RESULT	1	PLN
ATT22F8/c	ATT22F8	89035 bp
LOCUS	Arabidopsis thaliana DNA chromosome 4, BAC clone T22F8 (E55A project).	27-MAY-1999
DEFINITION	Arabidopsis thaliana DNA chromosome 4, BAC clone T22F8 (E55A project).	
ACCESSION	AL050351	
VERSION	AL050351.1	GI:4914422
KEYWORDS	thale cress.	
SOURCE	Arabidopsis thaliana	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.	
REFERENCE	1 (bases 1 to 89035)	
AUTHORS	Bevan, M., Zimmermann, W., Gruenisen, A., Mambutt, R., Bancroft, I., Hewes, R.W., Meyer, K.F.X. and Schellier, C.	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 89035)	
AUTHORS	EU Arabidopsis sequencing, project.	
TITLE	Direct Submission	



JOURNAL Submitted (27-MAR-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/chal/>.

FEATURES  
source location/Qualifiers  
1. .89035  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
1. .7016  
/note="Overlap to BAC F19H22: please refer to EMBL acc. AL035679 for analysis and annotation"  
gene complement(5175. .7811)  
/gene="T22F8.10"  
5175. .7811  
/gene="T22F8.10"  
complement(5175. .7811)  
/note="similarity to receptor-like protein kinase, Catharanthus roseus, Z73295"  
/codon\_start=1  
/product="putative receptor-like protein kinase"  
/protein\_id="CAB43626.1"  
/db\_xref="GI:4914423"  
/translation="MEIRKKRNIFVLYVIDSSKRSALLAILFLSPSASAVAA AVGATGKRPADDLIDGSSKSTPGRVKFSQETIQTLEAKEDIQVSPSDKY ASPIYLARIFREARYKFLTRPGMAWRULFLAPNDKEDLQATFSLVEKYL HNFKISNNNDQAAVQKLEYLVNMTDAQFALFRPKSSAAFINALEVSADELI SDGTALFVIGSGSDYAYOSVRYNNGPLIMPDIDKEKLEKDENLAKD VKTPSAIKYPEVTPPLIAPQTVAVAVEMANSLTIDNPNVSNMPEPSPNYIRL HFCDIVSKSLNDIENYINIKTAISGDLSTVAGNLAPYKDIYVATLKGPELOY QICPMGEDTGTKNAILNGVEVLKMSNSVNSLDGERGVDTGKRGHGMVATAGVMM FGAFILGAMAVYKMKRKPQDMQKRSFSSWLLPIHAGDSTFWTSKGSOKSFYVSTL GILRVYFSISELOEATKNEASQIIGVGCGFVNYIGTLDGTRKVAVRGNPQSEQITE FQETIOMLSLRHRLVSLIGCDENSEMILYVEPMSGPFDRHLGKRLAPLTWKOR LEICIGSARGIHLHTGAOGIIRHDKVSTNILDALVAVARDGLSKDVAFGONHV STAVKSGSYLDEPERFRROQLFDKSDVSEFVYLEALCARPAINPOLPREVNIAYE AMQKRRGLLEKTIIDPLAGTINPESKKEFAAKCLEDYGVDRPTMGDVLNMLEYA LQIQEAFITQGAETENAKPDVYTPGVSVSDPSITPSTVTTNEATVPVPAKVEENS GTAVDEHSGTAMFTGFANLNGR"  
exon complement(5175. .7811)  
/gene="T22F8.10"  
/number=1  
8270. .10413  
/gene="T22F8.20"  
complement(8270. .8340)  
/gene="T22F8.20"  
/number=1  
complement(join(8270. .8340,8460. .8571,8701. .8790, 9059. .9125,9296. .9420,9594. .9643,9732. .9826,10010. .10088, 10242. .10413))  
/gene="T22F8.20"  
complement(join(8270. .8340,8460. .8571,8701. .8790, 9059. .9125,9296. .9420,9594. .9643,9732. .9826,10010. .10088, 10242. .10413))  
/gene="T22F8.20"  
/note="similarity to Mono-phosphatase, Streptomyces anulatus, X92429  
contains inositol monophosphatase family signatures [WVLDPIDGTRKSFRT][VLDPIDGTRKSFRT]"  
CDS /codon\_start=1  
/product="inositol monophosphatase-like protein"  
/protein\_id="CAB43627.1"  
/db\_xref="GI:4914424"  
/translation="MASNRPNISNESPSLPTDELDRFAVGNALADASGEVIRKY FRKKFSLVDKDMSPVTLADQMAEEAMWSITFQNPISAIYIGEEKGMCKGKESADVAV VLDPIDGTRKSFRTTKPVGTLIALLYKRPILGLIDPILKERWIGMGRRTKLNGED

ISTRSCPILSQAIVLYTTPHLPSEBAEAKYRVRDKYKVPICGDCYAYALLASGFD IVIESGLKPYDFLALVPVIEAGGTTIDWTKRRLMEASSSAVTSFNVAAGDSDIH QOALSSLMH"  
intron complement(8341. .8459)  
/number=1  
complement(8460. .8571)  
/gene="T22F8.20"  
/number=2  
complement(8572. .8700)  
/number=2  
complement(8701. .8790)  
/gene="T22F8.20"  
/number=3  
complement(8791. .9058)  
/number=3  
complement(9059. .9125)  
/gene="T22F8.20"  
/number=4  
complement(9126. .9295)  
/number=4  
complement(9296. .9420)  
/gene="T22F8.20"  
/number=5  
complement(9421. .9593)  
/number=5  
complement(9594. .9643)  
/gene="T22F8.20"  
/number=6  
complement(9644. .9731)  
/number=6  
complement(9732. .9826)  
/gene="T22F8.20"  
/number=7  
complement(9827. .10009)  
/number=7  
complement(10010. .10088)  
/gene="T22F8.20"  
/number=8  
complement(10089. .10241)  
/number=8  
complement(10242. .10413)  
/gene="T22F8.20"  
/number=9  
complement(join(10875. .11060,11267. .11536))  
/gene="T22F8.30"  
10875. .11536  
/gene="T22F8.30"  
complement(join(10875. .11060,11267. .11536))  
/gene="T22F8.30"  
/note="similarity to dehydrin, glycine max, AF004807"  
intron /codon\_start=1  
/product="putative protein"  
/protein\_id="CAB43628.1"  
/db\_xref="GI:4914425"  
/translation="MADLKDRGNSPIYLTDAHSGPQALMDFGNAMHLTGVATTPHL KESSYTGPHPTATVTTNTNPHHQAIPISYSDPLQDDDLMPFGTSSYEENGBGVKRT NITDETCKSLGVDPKPSAATVSGSSGVHEKKGFKIKIKELSGHNDL"  
exon complement(10875. .11060)  
/gene="T22F8.30"  
/number=1  
complement(11061. .11266)  
/number=1  
complement(11267. .11536)  
/gene="T22F8.30"  
/number=2  
complement(12329. .13285)  
/gene="T22F8.40"  
/number=1  
12329. .13974  
/gene="T22F8.40"  
complement(join(12329. .13285,13369. .13415,13689. .13974))  
/gene="T22F8.40"  
complement(join(12329. .13285,13369. .13415,13689. .13974))  
CDS



```
/gene="T22F8.40"
/note="contains EST gb:T42243, AA58588"
/codon_start=1
/product="putative protein"
/protein_id="CA33629.1"
/db_xref="GI:4914426"
/translation="MGNACCAVAAARDKRVVNSAGENLQSRNIRHSPSFRWNRGR
VAGEETSLMSLSDGISRNDSSEIKFEASVSSGSPDSEFTQLOKSPASDLSPRN
SSMNVFEOKENVSTESAPSYLSIPAOLSLASQSSPFTQSSQSLHAPASSTL
KITQHPRLSKOVSDQITGVONSLSSASATEERAGCPPTRRDSSQSPBSMISQAPSEM
MSSSRNKAISLSDNIDNDFGLQDRKIDIHGNRMSKHQHTGCACSRPISEKLSMSO
KIFMTNELSVSALACGHVHGECEQMPEIDKPPSCPICTMGERRKAKLSEKALK
VEMDLKARHNKRLRNVRVLDSEDCDFVMEFDSHRTAAASRPRLVSSSAKSYSAK
PELARHFSFGSRNYSKPKENLPVKKKGFMTKSKT"
complement(13286..13368)
Intron
/number=1
complement(13369..13415)
exon
/number=2
/gene="T22F8.40"
/number=2
complement(13416..13688)
Intron
/number=2
complement(13689..13974)
exon
/number=3
complement(13975..16441,16515..16595,16713..16885,
16990..17148,17278..17407,17493..17658,17826..17908,
18062..18121,18346..18432)
/gene="T22F8.50"
16343..18432
/gene="T22F8.50"
complement(16343..16441,16515..16595,16713..16885,
16990..17148,17278..17407,17493..17658,17826..17908,
18062..18121,18346..18432)
/gene="T22F8.50"
/note="similarity to CAJ1 protein, Saccharomyces
cerevisiae, PIR2:S48085"
Query Match 44.2%; Score 54.4; DB 8; Length 89035;
Best Local Similarity 80.0%; Pred. No. 2.9e-08;
Matches 64; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 caggatctgtaactgtagatgccttaccacccgtgacctgaatgagcgaatgagag 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2099 CAGGTCTCGAAGTGTGAGATGCCGTATACCCAGATGACTTGTGTGATGCGAGCA 2040
QY 61 ctgcaactgactgtaagtt 80
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2039 GTCTTCTGAGTGTGAGTGT 2020
RESULT 2
ATF19H22/c ATF19H22 100469 bp DNA PLN 27-AUG-1999
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone F19H22 (ESSA
DEFINITION project).
ACCESSION AL035679
VERSION AL035679.1 GI:4539309
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 100469)
Bevan, M., Murphy, G., Ridley, P., Hudson, S., Bancroft, I., Mewes, H.W.,
Mayer, K.F.X. and Scheller, C.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 100469)
REFERENCE EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schnell@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
```

```
COMMENT
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@brc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosome 4 can be
viewed at: http://webserv.mips.biochem.mpg.de/proj/thal/.
FEATURES
source
misc_feature
gene
gene
gene
exon
CDS
complement(join(20670..21072,21414..21540,21709..22405,
22506..22616))
/gene="F19H22.10"
20670..22616
/gene="F19H22.10"
complement(20670..21072)
/gene="F19H22.10"
/number=1
complement(join(20670..21072,21414..21540,21709..22405,
22506..22616))
/gene="F19H22.10"
/note="similarity to T32F12.7 Arabidopsis thaliana
chromosome II BAC T32F12, PID:g3608133"
/codon_start=1
/product="putative protein"
/protein_id="CA33881.1"
/db_xref="GI:4539310"
/translation="MRDINSAIVSARSKRMESGGVNGRYKPPYKGTOSVNMPPKE
QNRALVNMKKIITSLAERDPAKRENEAYATREALASDEALBQDKALSDNKLIM
ETESALNALRYRENLDNLYLSCKRRGSGRFTEESHLPNPSITIPPEAATRPK
RKRESQKGMGDELNPVAPSPGKSKRMDSDNVTFEDENMPVPMCTCTGARQC
YKNGNGMGSCTTTLSEYPLPQPNKRHSRYGGRKMSGVSRLSLADLTIRIK
SRMKDEVRIVIVTVDSVYKROISSTVLTMFCCFIEATFEFGIARAFETWLNKHS
GQWITVYALSLIILPILIMIGIYMTFSSSGMSOVAFPWVEEOCYDGSAREGN
KETKDNCSLAIYVNSADQKCKMEKIFENKTKAKLRVYFPLHSYRSCRT"
complement(21073..21413)
/number=1
complement(21414..21540)
/gene="F19H22.10"
/number=2
complement(21541..21708)
/number=2
complement(21709..22405)
/gene="F19H22.10"
/number=3
complement(22406..22505)
/number=3
complement(22506..22616)
/gene="F19H22.10"
/number=4
23475..23550
/gene="F19H22.20"
/number=1
23475..24998
/gene="F19H22.20"
join(23475..23550,24243..24528,24866..24998)
/gene="F19H22.20"
/BC_number="3.6.1.35"
/note="one AA difference to PIR2:S60132
contains EST gb:T45164, R90510, T46102, F14129"
/codon_start=1
/product="H+-transporting ATPase 16k chain p2, vacuolar"
/protein_id="CA338812.1"
/db_xref="GI:4539311"
/translation="MSTFSGDETAPFFGFLGAALVFSCKMANGTAKSGVAVSMG
VWRPELVNKSIVPVWAGVGLIIVLITGINKPKASYLLFDGVNHLSSGLACG
LGLSGNAIGIVDAGVRANQOQPLFVGMILILIFAEALALYGLIVGIIILSRAGQ
SRAE"
```



```

intron      23551..24242
            /gene="F19H22.20"
            /number=1
exon        24243..24528
            /gene="F19H22.20"
            /number=2
intron      24529..24865
            /gene="F19H22.20"
            /number=2
exon        24866..24998
            /gene="F19H22.20"
            /number=3
gene         26052..27761
            /gene="F19H22.30"
            /note="ESTs GB:227052, Z29899, H76778, T76313, H76773,
            T76734 are in 3' untranslated region
            EST T42894 covers Intron in 5' untranslated region
            strong similarity to ubiquitin fusion-degradation protein
            - Mus musculus, PID:g1654348
            contains EST gb:AA395284, T42894"
            /codon_start=1
            /product="putative ubiquitin-dependent proteolytic
            protein"
            /protein_id="CAB38813.1"
            /db_xref="GI:4539312"
            /translation="MFYDGYAVHGTTPOTYRCYPSFIDKPOIESGDKIMPSALD
            RLASLIDYPMLELRNASTDSFSHGYLEFAEGVITYIPYMMONILLQSGDVRV
            RNVLPKCTYKLOPHRTPLDIANPKAILELALNYSCLTYGDSIMPYNNKKTFID
            IVEAKPSNGISITIEDCEVDPAFPLDYKEPERKVPAPAPKKGAKKEVDVAEKEFN
            PFTSGRGLDRLPSLEPQPGANNSGQSHPIYAVASSSSGSEKATQONRGLKVGSNVE
            RSTETTKVGAGSKRKOEEAEKEAKFQAFSGKKYSLNG"
            /number=1
exon        26052..26054
            /gene="F19H22.30"
            /number=1
intron      26055..26137
            /gene="F19H22.30"
            /number=1
exon        26138..26215
            /gene="F19H22.30"
            /number=2
intron      26216..26339
            /gene="F19H22.30"
            /number=2
exon        26340..26364
            /gene="F19H22.30"
            /number=3
intron      26365..26464
            /gene="F19H22.30"
            /number=3
exon        26465..26497
            /gene="F19H22.30"
            /number=4
intron      26498..26588
            /gene="F19H22.30"
            /number=4
exon        26589..26710
            /gene="F19H22.30"
            /number=5
intron      26711..26792
            /gene="F19H22.30"
            /number=5
exon        26793..26923
            /gene="F19H22.30"
            /number=6
intron      26924..27037
            /gene="F19H22.30"
            /number=6
exon        27038..27258
            /gene="F19H22.30"

```

```

intron      /number=7
            27259..27333
            /gene="F19H22.30"
            /number=7
exon        27334..27572
            /gene="F19H22.30"
            /number=8
intron      27573..27665
            /gene="F19H22.30"
            /number=8
exon        27666..27761
            /gene="F19H22.30"
            /number=9
repeat_region 27950..27971
            /note="22bp T tandem repeat"
            28993..30105
            /gene="F19H22.40"
            /number=1
gene         28993..30105
            /gene="F19H22.40"
            /number=1
CDS         28993..30105
            /gene="F19H22.40"
            /note="similarity to more than 30 predicted proteins,
            Arabidopsis
            contains EST gb:AB015106"
            /codon_start=1
            /product="putative protein"
            /protein_id="CAB38814.1"
            /db_xref="GI:4539313"
            /translation="MSPPIIRAPKKOPPEPCLISLPEIIVDIYARVRCYPTL
            SGVRSRFSIYASPEIYRBRFCCTECTLYAISKQTSIHPTLCRPNGQFSG
            TTASHRLVHPTLPMPMHSYIGISNITVWKGFCMKRTTSVSLDCRTHRAQTL
            PMPRAVAVPTVELIDRKIRIVIGSDTLSPKSPSRIMMVDTDTMQLNRPDWEA
            44.2%; Score 54.4; DB 8; Length 100469;
            Best Local Similarity 40.0%; Pred. No. 2.9e-08;
            Matches 64; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
            QY 1 caggatctgtaaatgtagatgcctacaacccatgacctaatgagcgcaatgcgaagg 60
            ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
            Db 95552 CAGGTTCTGCAGAGTCGAGATGCCGTATACCAGATGACTTGATGATCGCAGAGGA 95493
            QY 61 ctgcactgactgtaagttt 80
            || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
            Db 95492 GTTCTCTGAGTGTGATGCT 95473

RESULT 3
LOCUS ATCHRIY90 198987 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 90.
ACCESSION AL161594
VERSION AL161594.2 GI:7270858
KEYWORDS
SOURCE
ORGANISM
    thale cress.
    Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1779)
AUTHORS Robben,J., Braeken,M., Glymonprez,B., Volckaert,G., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
            Unpublished
JOURNAL 2 (bases 131465 to 198987)
AUTHORS Zimmermann,W., Grueniseisen,A., Wambutt,R., Kalicki,J., Wohlmann,P.,
            Smith,A., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
            Unpublished
REFERENCE 3 (bases 1 to 55749; 38011 to 138479)
AUTHORS Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
            Unpublished
JOURNAL 4 (bases 1 to 198987)
AUTHORS EU Arabidopsis sequencing project.

```



TITLE	Direct Submission
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemche@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ARCHIRV91 at the 5' end and an overlap with ARCHIRV91 at the 3' end.
FEATURES	Location/Qualifiers 1..198987
source	/organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="ATg38760" 5956..9755 /gene="ATg38760" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="ATg38760" /note="similarity to KIAA0169 gene, Homo sapiens, D79991" /codon_start=1 /product="putative protein" /protein_id="CAB80539.1" /db_xref="GI:7270859" /translation="MDVEKIDMINDIEIOSYILVERSEOEYGTDSVAOELTQEFIDMSFPHINOCFFVMWSEAFLLTLISFYSNVLEEDASVLHTAPVTKYKAYSHTCICSYVARRESSIKEEAVKLISDLERROSSVLEDLSSCPKNMGITLSGSYFKLAVSAOHASACROIOLOMLIETLEDMENLOWHDGVPFRSGCVSIIVDVENMATLSLTSPNVNAGPLVALMANVFCLTSLISLGKESPTIMDIHVSYVHOAEASISYELETIOSNLNDPGLPSIGHSRYRTTSAPISYEINTLOLEDGLELILDILSKYQDEESTCCQCFWDRKSFPVDGPIRCLFLDESEPPFSAEFILSLSESWPACVYNPLDKSVYSTLEDITSDSPADASQLVETSRLHPGLEGVIPSTNGRIIVYSNNYLVRXEVSTLSIIIVILIRANKLYIGNNRFAVTELLRRMTFNKAVCSSLINISHPFYVESYVNGKMESDVRYVDICTNSVRSLEFDSCGAVMAMAILIAKLRLW"
gene	complement(5956..6064) /gene="ATg38760" /number=1 complement(6065..6140) /number=1 complement(6141..6226) /gene="ATg38760" /number=2 complement(6227..6419) /number=2 complement(6420..6539) /gene="ATg38760" /number=3 complement(6540..6641) /number=3 complement(6642..6843) /gene="ATg38760" /number=4 complement(6844..7031) /number=4 complement(7032..7189) /gene="ATg38760" /number=5 complement(7190..7274) /number=5 complement(7275..7334) /gene="ATg38760" /number=6
CDS	
exon	
intron	
gene	
FEATURES	Location/Qualifiers 1..198987
source	/organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="ATg38760" 5956..9755 /gene="ATg38760" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="ATg38760" /note="similarity to KIAA0169 gene, Homo saplens, D79991" /codon_start=1 /product="putative protein" /protein_id="CAB80539.1" /db_xref="GI:7270859" /translation="MDVEKIDMINDIEIOSYILVERSEOEYGTDSVAOELTQEFIDMSFPHINOCFFVMWSEAFLLTLISFYSNVLEEDASVLHTAPVTKYKAYSHTCICSYVARRESSIKEEAVKLISDLERROSSVLEDLSSCPKNMGITLSGSYFKLAVSAOHASACROIOLOMLIETLEDMENLOWHDGVPFRSGCVSIIVDVENMATLSLTSPNVNAGPLVALMANVFCLTSLISLGKESPTIMDIHVSYVHOAEASISYELETIOSNLNDPGLPSIGHSRYRTTSAPISYEINTLOLEDGLELILDILSKYQDEESTCCQCFWDRKSFPVDGPIRCLFLDESEPPFSAEFILSLSESWPACVYNPLDKSVYSTLEDITSDSPADASQLVETSRLHPGLEGVIPSTNGRIIVYSNNYLVRXEVSTLSIIIVILIRANKLYIGNNRFAVTELLRRMTFNKAVCSSLINISHPFYVESYVNGKMESDVRYVDICTNSVRSLEFDSCGAVMAMAILIAKLRLW"
gene	complement(5956..6064) /gene="ATg38760" /number=1 complement(6065..6140) /number=1 complement(6141..6226) /gene="ATg38760" /number=2 complement(6227..6419) /number=2 complement(6420..6539) /gene="ATg38760" /number=3 complement(6540..6641) /number=3 complement(6642..6843) /gene="ATg38760" /number=4 complement(6844..7031) /number=4 complement(7032..7189) /gene="ATg38760" /number=5 complement(7190..7274) /number=5 complement(7275..7334) /gene="ATg38760" /number=6
CDS	
exon	
intron	
gene	
FEATURES	Location/Qualifiers 1..198987
source	/organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="ATg38760" 5956..9755 /gene="ATg38760" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="ATg38760" /note="similarity to KIAA0169 gene, Homo saplens, D79991" /codon_start=1 /product="putative protein" /protein_id="CAB80539.1" /db_xref="GI:7270859" /translation="MDVEKIDMINDIEIOSYILVERSEOEYGTDSVAOELTQEFIDMSFPHINOCFFVMWSEAFLLTLISFYSNVLEEDASVLHTAPVTKYKAYSHTCICSYVARRESSIKEEAVKLISDLERROSSVLEDLSSCPKNMGITLSGSYFKLAVSAOHASACROIOLOMLIETLEDMENLOWHDGVPFRSGCVSIIVDVENMATLSLTSPNVNAGPLVALMANVFCLTSLISLGKESPTIMDIHVSYVHOAEASISYELETIOSNLNDPGLPSIGHSRYRTTSAPISYEINTLOLEDGLELILDILSKYQDEESTCCQCFWDRKSFPVDGPIRCLFLDESEPPFSAEFILSLSESWPACVYNPLDKSVYSTLEDITSDSPADASQLVETSRLHPGLEGVIPSTNGRIIVYSNNYLVRXEVSTLSIIIVILIRANKLYIGNNRFAVTELLRRMTFNKAVCSSLINISHPFYVESYVNGKMESDVRYVDICTNSVRSLEFDSCGAVMAMAILIAKLRLW"
gene	complement(5956..6064) /gene="ATg38760" /number=1 complement(6065..6140) /number=1 complement(6141..6226) /gene="ATg38760" /number=2 complement(6227..6419) /number=2 complement(6420..6539) /gene="ATg38760" /number=3 complement(6540..6641) /number=3 complement(6642..6843) /gene="ATg38760" /number=4 complement(6844..7031) /number=4 complement(7032..7189) /gene="ATg38760" /number=5 complement(7190..7274) /number=5 complement(7275..7334) /gene="ATg38760" /number=6
CDS	
exon	
intron	
gene	
FEATURES	Location/Qualifiers 1..198987
source	/organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="ATg38760" 5956..9755 /gene="ATg38760" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7



[illegible]



32529. .326071)  
/gene="AT4g21990"  
/note="complements functionally an Escherichia coli mutant defective in PAPS reductase activity  
Contains Prokaryotic membrane lipoprotein lipid attachment site AA276-286  
contains EST gb:234086, Z35366"  
/codon\_start=1  
/product="PRH26 protein"  
/protein\_id="CAI8102.1"  
/db\_xref="GI:2961344"  
/db\_xref="SPTREMBL:O48887"  
/translation="MALAINVSSSSSAISSSPSSDLKTKIGSLRLNRTVNSAASLSLGSKRSVALNVSITKESIVASEVTEKLDVDEDEELAKRLKLENSPSEIMDKLEKFGNDIAIAFSGARDVALIEFAHLITGRPYRSLDGRINDEYRLIDPYEKYGRIRIYMPDAVEOVALYRNKGLPSEFYEDGHQECRIKAYRPLRALKGLAAMITGQRKQSGFSTSEIIVOVDPVEFSLDGVGSLVKMNVANVEGNDVNEFLRTMDPVNLTGAAVSADYISGCEPTRAVLPGQHERGRWMBDPAKKEGLAKGNIKEMTNOANAVNGHVSADYIFNSENVNVLNROGILENLMKLENKEAMIVLVLPWCFFCQAMASDELDKLGSGGVAKAFRADGQDKPAKKEQLGSPPTLTVPKNSSRIPKYSBKRDVSLTSEFLIVVR"  
complement(30918. .31944)  
/gene="AT4g21990"  
/number=1  
complement(31945. .32029)  
/number=1  
complement(32030. .32165)  
/gene="AT4g21990"  
/number=2  
complement(32166. .32251)  
/number=2  
complement(32252. .32386)  
/gene="AT4g21990"  
/number=3  
complement(32387. .32528)  
/number=3  
complement(32529. .32607)  
/gene="AT4g21990"  
/number=4  
complement(33444. .33836)  
/gene="AT4g22000"  
/number=1  
complement(33444. .33836)  
complement(33444. .33836)  
/gene="AT4g22000"  
/number=1  
complement(33444. .33836)  
/gene="AT4g22000"  
33444. .33836  
/gene="AT4g22000"  
complement(33444. .33836)  
/gene="AT4g22000"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAI8103.1"  
/db\_xref="GI:2961345"  
/db\_xref="SPTREMBL:O65448"  
/translation="MAGKAAEAVATVTFQHPMPRAKLDKRYTELTGQWGYWEGAVKPLGSARRAALRKVELTNGDMPDYDERKAMRTKKRGHCDRIASAKRENTAKMLKMPQMLDLYKKRWEKKMKKEBEKAKEDK"  
37063. 37197  
/gene="AT4g22010"  
/number=1  
37063. .40097  
/gene="AT4g22010"  
join(37063. 37197,37310. 37419,38362. .38632,38736. .39184,39290. 39658,39806. .40097)  
/gene="AT4g22010"  
/note="strong similarity to peclinesterase, Lycopersicon esculentum, PATX:E312172  
contains EST gb:AI994449.1, Z26709, T43750, Z26220, AI099778"  
/codon\_start=1  
/product="pectinesterase like protein"  
/protein\_id="CAI8104.1"  
/db\_xref="GI:2961346"  
/db\_xref="SPTREMBL:O65449"

[illegible]



COMMENT	FEATURES
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> this fragment has an overlap with ATCHRIV55 at the 5' end and an overlap with ATCHRIV57 at the 3' end.	location/Qualifiers
	1. 196247
	/organism="Arabidopsis thaliana"
	/variety="Columbia"
	/db_xref="taxon:3702"
	/chromosome="4"
gene	5510. 11823
exon	/gene="AT4g21820"
	5510. 6217
	/gene="AT4g21820"
	/number=1
CDS	join(5510. 6217,6310. 6778,7223. 7449,7548. 7613. 7799. 7852,8220. 8336,8544. 8642,8699. 8879,9087. 9224, 9426. 9620,9702. 9982,10872. 11094,11161. 11425, 11469. 11624,11736. 11823)
	/note="similarity to calmodulin-binding protein SH1, Mus musculus, AF062378
	contains EST gb:AI97648.1"
	/codon_start=1
	/product="putative protein"
	/protein_id="CAB79137.1"
	/db_xref="GI:7269027"
	/translation="MDENPPCASPPAPRNPASSLIDSNFKTPRRTSVVNSNISKSPYHFFRTASKOTPKSSSNFRPRSMVHSYASRKYSTSSRRRLKAFELQOSOSRAELTKENLRSLASLTIVNLPLEPENCGCDPENSGVNLGKGRDSEALGNSKSVGDTMMRSPKRLRLNWCWGEKRSEIDSLGSKYSTLRESLDYCSLDLDRQMPHLYISGCKEIPDMTFRVSKNIDEGRIKMKPOCELVDFGKKAIKALMSYNQVRLGLYIIFEGDSFLDSEVNSDOEMAFKAVISKQFSDHGLARSYYVKNKWEGLRPGY EALGNVILKRLILVIVIDRAKSOSCLSIKYGIDGIDGSPIMFSEKSKISSHOLICGSEEDLMRVILKQKRVVPSDNKRKRLALQILKQAGVSLKDEGCMITGEVADGDRELTISLWNIFVHQLPLINGRLLFEETIKYQVGEQNNQITMSTPLELWNWIDPQGQEGPQSVASNTDYHDVQNFILSQKLALSLSPFEGDLNHNVAVNSQVYIILLAFLSKGLIVKENSNDFTLAGLTNSVSOEQLNFKLGCSCQVQEKRSIRISCTSSSAVTIEEPRENEDATKRFQAIKAMMODANQNOISVGCANSHHTLOGSLRGT DSORAVILLAEIAVITIOSNIRGLHARRKPRMKRAICFLOAAVRTWISVKNIQVVEKFNVEEVTILHSEBSANLKPARYKFTYDRSEFTLRKSVSYTQKAVRRHDSNLHIELKALRIQLAKRSYKEKVISITIOSIVGWITRRNRYKESILIOFOVSTPYTISQSHWGYLTRKASKAQVLDLRVMQTSANIDKKRLINKLALSSELLSMKNVNIHLICETLESATKYSIDKCEELVAAGAIEKLTLIRASRSIPDOQVSKHALSTIGHLARY POMADELINTKSGIOTIMELIRSSQVEFLSPKNEEAFVFIASDVLRKICSSHGKVEAVRKLPALVKRLHVLVEELLRRANIEKRNKVGQTGKEKSERLKEALEIVKLITRR"
intron	6218. 6309
	/gene="AT4g21820"
	/number=1
exon	6310. 6778
	/gene="AT4g21820"
	/number=2
intron	6779. 7222
	/gene="AT4g21820"
	/number=2
exon	7223. 7449
	/gene="AT4g21820"
	/number=3
intron	7450. 7547
	/gene="AT4g21820"
	/number=3
exon	7548. 7613
	/gene="AT4g21820"
	/number=4
intron	7614. 7798
	/gene="AT4g21820"
	/number=4
exon	7799. 7852
	/gene="AT4g21820"
	/number=5
intron	7853. 8219
	/codon_start=1
	/gene="AT4g21820"
	/number=5
exon	8220. 8336
	/gene="AT4g21820"
	/number=6
intron	8337. 8543
	/gene="AT4g21820"
	/number=6
exon	8544. 8642
	/gene="AT4g21820"
	/number=7
intron	8643. 8698
	/gene="AT4g21820"
	/number=7
exon	8699. 8879
	/gene="AT4g21820"
	/number=8
intron	8880. 9086
	/gene="AT4g21820"
	/number=8
exon	9087. 9224
	/gene="AT4g21820"
	/number=9
intron	9225. 9425
	/gene="AT4g21820"
	/number=9
exon	9426. 9620
	/gene="AT4g21820"
	/number=10
intron	9621. 9701
	/gene="AT4g21820"
	/number=10
exon	9702. 9982
	/gene="AT4g21820"
	/number=11
intron	9983. 10871
	/gene="AT4g21820"
	/number=11
exon	10872. 11094
	/gene="AT4g21820"
	/number=12
intron	11095. 11160
	/gene="AT4g21820"
	/number=12
exon	11161. 11425
	/gene="AT4g21820"
	/number=13
intron	11426. 11468
	/gene="AT4g21820"
	/number=13
exon	11469. 11624
	/gene="AT4g21820"
	/number=14
intron	11625. 11735
	/gene="AT4g21820"
	/number=14
exon	11736. 11823
	/gene="AT4g21820"
	/number=15
gene	12517. 13199
	/gene="AT4g21830"
	complement(12517. 12675)
exon	/gene="AT4g21830"
	/number=1
gene	complement(join(12517. 12675,12834. 12990,13090. 13199))
	/gene="AT4g21830"
	/note="strong similarity to CGI-131 protein, Homo sapiens, AF151889
	contains EST gb:N38012, AA713335, AI996517.1, AA712450, H76478"
	/codon_start=1



```

/product="putative protein"
/protein_id="CAB79138.1"
/db_xref="GI:7269028"
/translation="MTAAVPATGSGFROKDEEMRAVLSPROFVLPLKGNDRKGEF
TKREEGTYSACAGCTALYKSTTKFSGCGWPPFALIREALVOTPEAGRRMETICA
VCDGHLGHVFKGEGYSTPDRHRCVNSVSLKSSAGSSQ"
complement(12676..12833)
/number=1
Intron
complement(12834..12990)
/gene="AT4g21830"
/number=2
Intron
complement(12991..13089)
/number=2
exon
complement(13090..13199)
/gene="AT4g21830"
/number=3
gene
15183..16130
/gene="AT4g21840"
complement(15183..15341)
/gene="AT4g21840"
/number=1
complement(join(15183..15341,15764..15920,16015..16130))
/gene="AT4g21840"
complement(join(15183..15341,15764..15920,16015..16130))

Query Match
Best Local Similarity 79.7%; Pred. No. 6.7e-08; Length 196247;
Matches 63; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 aggtattgtaaatgtgagtccttacaacctgatacctaagcgcaatgcgagggc 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156910 AGGTACTGCAAAATGTAAGTCCCTTACATCCAGACGATCTCATGCTGTAAGGC 156851
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 62 tgcactactggtgaagttt 80
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156850 TGCAAAGACTGTAATATGT 156832

RESULT 6
AP003753
LOCUS
DEFINITION
Oryza sativa chromosome 7 clone OJ1339_B08, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION
AP003753
VERSION
AP003753.1 GI:14422471
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1339_B08.
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (sites)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone:OJ1339_B08
Published Only in Database (2001) In press
2 (bases 1 to 119807)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/,
Tel:81-298-387441, Fax:81-298-387468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.

```

```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
source
1..119807
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="7"
/clone="OJ1339_B08"

BASE COUNT 34992 a 23815 c 24131 g 36815 t 54 others
ORIGIN

Query Match
Best Local Similarity 42.6%; Score 52.4; DB 2; Length 119807;
Matches 62; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 caagttattgtaaatgtgagtccttacaacctgatacctaagcgcaatgcgaggg 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109849 CAGGTTCTGCAAGTGTGAGTGCATACATCTGATGACCTTATGATCAATGAGAGA 109908
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 ctgcactgactggtgaagt 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109909 ATGCTCTGATGTGTGAGT 109926

RESULT 7
AF277453
LOCUS
DEFINITION
Arabidopsis thaliana putative PHD finger transcription factor
(SHL1) mRNA, complete cds.
ACCESSION
AF277453
VERSION
AF277453.1 GI:10644644
KEYWORDS
thale cress.
SOURCE
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1062)
Oh,M.-H., Torisky,R.S., Braam,J., Altman,T. and Clouse,S.D.
PHD Finger Dependent Binding of SHL1 to a Specific Promoter Region
of Arabidopsis TCH4 In vitro
Unpublished
2 (bases 1 to 1062)
Clouse,S.D. and Oh,M.-H.
Direct Submission
Submitted (13-JUN-2000) Horticultural science, North Carolina State
University, 228 Kilgore Hall, Raleigh, NC 27695-7609, USA
Location/Qualifiers
1..1062
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="IV"
/notes="similar to BAC clone F19H22 in GenBank Accession
Number AL035679"
1..1062
/gene="SHL1"
58..744
/gene="SHL1"
/codon_start=1
/product="putative PHD finger transcription factor"
/protein_id="AA021353.1"
/db_xref="GI:10644645"
/translation="MPKQAPRKQSLKYLKHNKSIQEGDVALMRSSGPKPSYVAR
VHSEFIDARSHAKVAVRWYRPEESIGRPGHGAKEVLSHPQFSADTLEGCK
VHSEFSTYKIDSVGNDPFCRFREYNTGTGAFDDRVTVCKCMPYNPDDIAVQCEC
SEPHNSCIGTITIEAKKPDNPFCEBSCPOOALHNSNSTSNMRDAKVNRSLSEVTK
SKMKHTRRP"

BASE COUNT 310 a 207 c 242 g 303 t
ORIGIN

```



Query Match	38.2%;	Score 47;	DB 8;	Length 1062;
Best Local Similarity	78.9%;	Pred. No. 9.1e-06;		
Matches	56;	Conservative	0;	Mismatches 15;
			Indels	0;
			Gaps	0;

QY	64	cactgactgct	74
Db	537	TTCTGAGTGGT	547
QY	4	gtatgttaaatgctagatgccccttaacacccctgactgactaaatgcgaatgcgagggct	63
Db	477	gTTCGCAAGTGTGAGATGCCGTAAACCCAGATGACTGTATGGTGCATATGCCAGAGATC	536

RESULT	8			
HVES43				
LOCUS				
DEFINITION	H.VES43	1058 bp	mRNA	
ACCESSION	X77575	(Dbg576)	ES43 mRNA.	PLN
VERSION	X77575.1	GI:482918		
KEYWORDS	ES43 gene.			
SOURCE	barley.			
ORGANISM	Hordeum vulgare			

REFERENCE AUTHORS TITLE JOURNAL REFERENCE	1 (bases 1 to 1058) Speulman, E. and Salami, F. A barley cDNA clone with homology to the DNA-binding domain of the steroid hormone receptors Plant Sci. 106, 91-98 (1995) 2 (bases 1 to 1058) Speulman, E.

FEATURES	Location/Qualifiers
source	1. .1058

BASE COUNT	237 a	286 c	283 g	252 t
ORIGIN				

Query Match	36.9%;	Score 45.4;	DB 8;	Length.1058;
Best Local Similarity	77.5%;	Pred. No. 3.3e-05;		
Matches	55;	Conservative	0;	Mismatches 16;
			Indels	0;
			Gaps	0;

Oy	4	gtaatgaatgttagagttgcctacaaacctgatgacttaattcgcaatcgaaggctg	63
Dd	539	gttttggcaagtgtagatgccattacaacctgacagacttttatcatccagtctgaggaatg	538
Oy	64	caatgactggt	74
Dd	599	cttcgattgggt	609

## RESULT 9

AB011445	AB011445	367 bp	mRNA	PLN	03-JUL-1999
LOCUS	Triticum aestivum WESR4		mRNA for zinc-finger motif, partial cds.		
DEFINITION	AB011445				
ACCESSION	AB011445.1	GI:5360896			
VERSION					
KEYWORDS	WESR4; zinc-finger motif				
SOURCE	Triticum aestivum (strain:Chinese Spring)				
					seedling root cDNA to

REFERENCE	1 (sites)
AUTHORS	Nemoto, Y., Kawakami, N. and Sasakuma, T.
TITLE	Isolation of novel early salt-responding genes from wheat ( <i>Triticum aestivum</i> L.) by differential display
JOURNAL	Theor. Appl. Genet. 98, 673-678 (1999)
REFERENCE	2 (bases 1 to 367)
AUTHORS	Nemoto, Y.
TITLE	Direct Submission
JOURNAL	Submitted (21-Feb-1998) to the DDBJ/EMBL/GenBank databases. Yase Nemoto, Yokohama City University, Kihara Institute for Biological Research; Maloka, Totsuka, Yokohama 244-0813, Japan (E-mail:nemoto@yokohama-cu.ac.jp, Tel:81-45-820-1902, Fax:81-45-820-1901)
FEATURES	
source	location/Qualifiers 1..367

FEATURES	source	location/Qualifiers
gene	CDS	1..367 /organism="Triticum aestivum" /strain="Chinese Spring" /db_xref="taxon:4565" /clone="WESR4" /dev_stage="seedling" /tissue_type="root" 1..217 /gene="WESR4" <1..217 /gene="WESR4" /note="salt-stress responding gene" /codon_start=2 /product="zinc-finger motif" /protein_id="BA82157.1" /db_xref="GI:5360897" /translation="PYNPDLIVOCGCKDWEHPNMGWITEAKKLDITFLCADCANE NGAKPSNYSYSSSPSSDSKVEPKRRK"
BASE COUNT		104 a 76 c 96 g 91 t

Query Match	34.08;	Score 41.8;	DB 8;	Length 367;
Best Local Similarity	80.38;	Pred. No. 0.00058;		
Matches 49;	Conservative	0;	Mismatches 12;	Indels 0;
				Gaps 0;

**Dy**      22 gacctaacacccctgtagaacctaattggcgcaatgcsaggcgtcacactgactgtaagtcttcc 81  
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
**Db**      1 gccgtacaaccagatgacctcatgttgccaatgccagggatgcgaagaactggttcacatccc 60

Qy	82 a 82
	—
Db	61 A 61

RESULT	10		
AX059527			
LOCUS	AX059527	44723 bp	DNA
DEFINITION	Sequence 260 from Patent WO0055325.		PAT
			17-JAN-2001

SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana  
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta  
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots,  
Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis.











PIKSLVSOLOQKTEFLKGFPPALQLIAPFRNIDQLDLVLDDSTSKTILMLKRLTHPKQ  
 VLSLQIHALSENLKLPVDPPLIYGLDNLADQSAEMDEVKDKVAYMYDLIMWSHQFO  
 KIDWPGGADTLPIISVPEKKOSKRGCGPRYSRKSKEVTLNLSDEKSRILISELOTVE  
 ELISNPMKLEKAKKTVRKRRTKLISSFPACSSISKRKIMEVPIQOSQSESTLDEH  
 VFGQPEVAAQVDPDDHLDKTDODSSDSIPLINTEITDDPHVFTPILOSSEKNDANEG  
 NNVYDITDVKDQANANEDVDSQMDASERVPTHSGLDKREHNSSELENNANETDVG  
 KKQDSDLRETAISHIDLPIDQSSSEQAKIMSELIMVPRKDVNLMSLETFNAAANIY  
 LCLSHIIFEVYNIQDTMEAAASHSLGFAKAKKVTLESAAVSQALSQTPPEVNF  
 DDMKAMAVIYSHPIHDTNSKPNKOLESSPTPEDDITDSDIGAHDLDNQOEGYVD  
 SDSPAREKREKPTLSAEVFLVAELLSKKTGSEYELPSMSKSEFALFENITSAAPNT  
 EHITSSGYLISNKFLLSLAKPTINWSTLMEVIVLSLKLATTLTNQARAPOPERA  
 NHLQGLKSFKAAMKSRVKSWEPMKQFTVGPTEWFADITLYLPMIDSKHWGLA  
 INLGVWSVEILDPNTDLVEDEVRREIEPVVIMPYLIORYCKPECSQNHGLOPFYWK  
 RLDGLTKNLRSGDCGPVAM"  
 /gene="AT4g03970"  
 8639. .8737  
 /number=1  
 /gene="AT4g03970"  
 8738. .8949  
 /number=2  
 /gene="AT4g03970"  
 8950. .8992  
 /number=2  
 /gene="AT4g03970"  
 8993. .9236  
 /number=3  
 /gene="AT4g03970"  
 9237. .9340  
 /number=3  
 /gene="AT4g03970"  
 9341. .9454  
 /number=4  
 /gene="AT4g03970"  
 9455. .9526  
 /number=4  
 /gene="AT4g03970"  
 9527. .9676  
 /number=5  
 /gene="AT4g03970"  
 9677. .10447  
 /number=5  
 /gene="AT4g03970"  
 10448. .10561  
 /number=6  
 /gene="AT4g03970"  
 10562. .10689  
 /number=6  
 /gene="AT4g03970"  
 10690. .10770  
 /number=7  
 /gene="AT4g03970"  
 10771. .10832  
 /number=7  
 /gene="AT4g03970"  
 10833. .10905  
 /number=8  
 /gene="AT4g03970"  
 10906. .10955  
 /number=8  
 /gene="AT4g03970"  
 10956. .11074  
 /number=9  
 /gene="AT4g03970"  
 11075. .11198  
 /number=9  
 /gene="AT4g03970"  
 11199. .11279  
 /number=10  
 /gene="AT4g03970"  
 11280. .11425  
 /number=10  
 /gene="AT4g03970"  
 11426. .11509  
 /number=10  
 /gene="AT4g03970"

/number=11  
 11510. .11625  
 /gene="AT4g03970"  
 /number=11  
 11626. .11679  
 /gene="AT4g03970"  
 /number=12  
 11680. .11792  
 /gene="AT4g03970"  
 /number=12  
 11793. .11986  
 /gene="AT4g03970"  
 /number=13  
 11987. .12068  
 /gene="AT4g03970"  
 /number=13  
 12069. .12153  
 /gene="AT4g03970"  
 /number=14  
 12154. .12237  
 /gene="AT4g03970"  
 /number=14  
 12238. .12735  
 /gene="AT4g03970"  
 /number=15  
 20580. .22438  
 /note="pseudogene"  
 26352. .28320  
 /gene="AT4g03980"  
 26352. .26944  
 /gene="AT4g03980"  
 /number=1  
 join(26352. .26944, 27053. .27955, 28074. .28320)  
 /gene="AT4g03980"  
 /note="similarity to similar to maize transposon 'mudr  
 mudra protein (GB:AL021710)"  
 /codon\_start=1  
 /product="putative transposon protein"  
 /protein\_id="CA877866.1"  
 /db\_xref="GI:7267154"  
 /translation="MFKVSFLFWQPRLTIRITIOACVSEMSSTIFDQYSSSAFSTGLSD  
 ADSLFLGKKEKDKDENVFTLRMFVAVKHSSEFHTVKSDLTRYVLACIDENCSMKLRATK  
 AGSESESYVIRKYVSHHSCDSLRLNYSRQAFARTLGLRISNHLGCKLPDGPOLIEI  
 FRDHGVGINYSKAMVQEHVVELARGLDDSPREVLPBGYKLMRKVYISDGAHLTFSK  
 KGTILGASADQGNENLYPSAFAIVDEPNASDMFLPKCLLNIIPDNDLVFVSDRAAS  
 IASGLSENTPLAHHCICTPHLOKNLTHFRGSSLLIPVYNLAASRVYTKTERDLSFMKIT  
 NSDKRLAQYLMEVDVWRKSRAYSPSNRNIMTNSLAESVNALIKQNRREPIVCLFESI  
 RSLMTRFMNERRESSQHPASAVTINVGKMKASDYDSTFWLVEVCQVQOHEFEVKGDTK  
 THLVINDKRTCTCCMEDIDKFPYAHGIAAKHINLKNMFEVDEHSTYRMROYSESI  
 HPNGDMEYWEIPDIVEVICLPSTRVPSGRKKKRRIPSVMEHGRSQPKKLKSCRC  
 GSGGHNKSTCVAAI"  
 26945. .27052  
 /gene="AT4g03980"  
 /number=1  
 27053. .27955  
 /gene="AT4g03980"  
 /number=2  
 27956. .28073  
 /gene="AT4g03980"  
 /number=2  
 28074. .28320  
 /gene="AT4g03980"  
 /number=3  
 31256. .32611  
 /gene="AT4g03990"  
 complement(join(31256. .31513, 31718. .31843, 31964. .32611))  
 /gene="AT4g03990"  
 complement(join(31256. .31513, 31718. .31843, 31964. .32611))  
 /gene="AT4g03990"  
 /note="similarity to"  
 /codon\_start=1

Query Match

28.0%; Score 34.4; DB 8; Length 198022;



Best Local Similarity 78.8%; Pred. No. 0.34; Matches 41; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2 aggtatgttaattgtagatgccttaaacacctgatgacctaatgycgcaat 53  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 169843 AGTACTACAAGTGAATGCCGTATTACTGATGATGATGATGAT 169894

RESULT 13  
 AL161773 154195 bp DNA PRI 17-JUL-2001  
 LOCUS Human DNA sequence from clone RP11-212E4 on chromosome 13, complete  
 DEFINITION sequence.

ACCESSION AL161773  
 VERSION AL161773.21 GI:14970659  
 KEYWORDS HTG.

SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 TITLE 1 (bases 1 to 154195)  
 JOURNAL

COMMENT Direct Submission  
 Submitted (17-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk  
 On Jul 19, 2001 this sequence version replaced gi:13751315.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP database can be found at  
 http://www.sanger.ac.uk/projects/C-elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr13  
 RP11-212E4 is from the library RPCR-11.1 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-212E4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true right end of clone RP11-212E4 is at 154195 in this sequence. The true left end of clone RP11-23116 is at 133478 in this sequence. The true right end of clone RP11-472K17 is at 2000 in this sequence.

FEATURES  
 source Location/Qualifiers

1..154195  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /clone="RP11-212E4"  
 /clone\_lib="RPCR-11.1"  
 346..646  
 repeat\_region  
 /note="Alusx repeat: matches 1..302 of consensus"  
 1158..1196  
 repeat\_region  
 /note="L2 repeat: matches 2693..2731 of consensus"  
 1415..1720  
 repeat\_region  
 /note="AluY repeat: matches 2..298 of consensus"

repeat\_region 3397..3622  
 /note="MER20 repeat: matches 1..218 of consensus"  
 repeat\_region 6907..7214  
 /note="Alusx repeat: matches 1..308 of consensus"  
 repeat\_region 9368..9663  
 /note="Alusq repeat: matches 1..295 of consensus"  
 repeat\_region 9761..9869  
 /note="MIR repeat: matches 120..244 of consensus"  
 repeat\_region 13836..13879  
 /note="L2 repeat: matches 2632..2692 of consensus"  
 repeat\_region 13883..13918  
 /note="9 copies 4 mer aggg 86% conserved"  
 repeat\_region 14560..14837  
 /note="Alusx repeat: matches 1..277 of consensus"  
 repeat\_region 15134..15378  
 /note="LIME3 repeat: matches 5895..6144 of consensus"  
 repeat\_region 15675..15791  
 /note="LIME3A repeat: matches 6034..6158 of consensus"  
 repeat\_region 17512..17569  
 /note="L2 repeat: matches 2632..2692 of consensus"  
 repeat\_region 18079..18368  
 /note="Alusx repeat: matches 1..290 of consensus"  
 repeat\_region 18409..19128  
 /note="12 copies 60 mer 59% conserved"  
 repeat\_region 18514..18678  
 /note="5 copies 33 mer 83% conserved"  
 repeat\_region 18676..19822  
 /note="37 copies 31 mer 56% conserved"  
 repeat\_region 18721..18756  
 /note="9 copies 4 mer atat 83% conserved"  
 repeat\_region 18831..19282  
 /note="14 copies 33 mer 89% conserved"  
 repeat\_region 19265..19534  
 /note="9 copies 30 mer 60% conserved"  
 repeat\_region 19267..19330  
 /note="32 copies 2 mer ta 78% conserved"  
 repeat\_region 19510..19689  
 /note="3 copies 60 mer 73% conserved"  
 repeat\_region 19553..19849  
 /note="9 copies 33 mer 85% conserved"  
 repeat\_region 19821..19872  
 /note="26 copies 2 mer ta 75% conserved"  
 repeat\_region 19823..19882  
 /note="3 copies 20 mer 85% conserved"  
 repeat\_region 19874..19925  
 /note="26 copies 2 mer ta 76% conserved"  
 repeat\_region 20002..20114  
 /note="Tigger4(Lombi) repeat: matches 1..110 of consensus"  
 repeat\_region 20176..20241  
 /note="3 copies 22 mer 89% conserved"  
 repeat\_region 20925..21135  
 /note="L2 repeat: matches 2490..2708 of consensus"  
 repeat\_region 21566..21983  
 /note="19 copies 22 mer 61% conserved"  
 repeat\_region 21571..21990  
 /note="14 copies 30 mer 63% conserved"  
 repeat\_region 21573..21980  
 /note="17 copies 24 mer 64% conserved"  
 repeat\_region 21589..21932  
 /note="8 copies 43 mer 73% conserved"  
 repeat\_region 21593..21840  
 /note="8 copies 31 mer 71% conserved"  
 repeat\_region 21600..21995  
 /note="11 copies 36 mer 62% conserved"  
 repeat\_region 21609..21992  
 /note="12 copies 32 mer 64% conserved"  
 repeat\_region 21668..21907  
 /note="12 copies 20 mer 77% conserved"  
 repeat\_region 21696..21995  
 /note="5 copies 60 mer 72% conserved"  
 repeat\_region 21723..21953  
 /note="7 copies 33 mer 69% conserved"  
 misc\_feature 21755



```

/note="random repeat: Forced join. Assembly confirmed by
restriction digest data"
repeat_region
24306. .24603
/note="AluX repeat: matches 6. .301 of consensus"
26130. .26485
/note="THEIB repeat: matches 1. .364 of consensus"
repeat_region
26684. .26812
/note="L1M3 repeat: matches 6014. .6128 of consensus"
repeat_region
26813. .27099
/note="AluI repeat: matches 1. .299 of consensus"
27100. .27150
/note="L1M3 repeat: matches 6128. .6181 of consensus"
repeat_region
29523. .29678
/note="AluJo repeat: matches 121. .276 of consensus"
repeat_region
30783. .30850
/note="L1 repeat: matches 4 mer cctt 67% conserved"
32023. .32423
/note="Cpg island"
/misc_feature
evidence=not_experimental
32679. .32982
/note="L1M4 repeat: matches 7670. .7974 of consensus"
repeat_region
33887. .33942
/note="L14 copies 4 mer tata 73% conserved"
repeat_region
33894. .33941
/note="24 copies 2 mer at 75% conserved"
repeat_region
36836. .37970
/note="L1M4 repeat: matches 6648. .7735 of consensus"
repeat_region
40099. .40219
/note="L2 repeat: matches 2617. .2748 of consensus"
repeat_region
41059. .42291
/note="L1M1 repeat: matches 5090. .6332 of consensus"
repeat_region
43625. .43723
/note="L1P3 repeat: matches 6050. .6150 of consensus"
repeat_region
44551. .44586
/note="9 copies 4 mer tcat 86% conserved"
repeat_region
48297. .48379
/note="MIR repeat: matches 123. .211 of consensus"
repeat_region
49534. .49665
/note="6 copies 22 mer 67% conserved"
repeat_region
49540. .49655
/note="29 copies 4 mer caca 64% conserved"
repeat_region
49666. .49737
/note="18 copies 4 mer acaa 66% conserved"
repeat_region
49669. .49736
/note="34 copies 2 mer ca 69% conserved"
repeat_region
50000. .50085
/note="2 copies 43 mer 95% conserved"
repeat_region
50923. .50998
/note="L2 repeat: matches 2592. .2671 of consensus"
repeat_region
51533. .51968
/note="M12CB repeat: matches 1. .495 of consensus"
repeat_region
54022. .54321
/note="AluG repeat: matches 3. .297 of consensus"
repeat_region
54797. .55092
/note="AluSg repeat: matches 1. .290 of consensus"
repeat_region
56788. .56865
/note="HY3 repeat: matches 1. .78 of consensus"
repeat_region
58501. .58818
/note="THEIC repeat: matches 58. .369 of consensus"
repeat_region
60826. .60955
/note="L2 repeat: matches 2605. .2750 of consensus"
repeat_region
61291. .61320
/note="15 copies 2 mer tt 90% conserved"
repeat_region
61468. .61770

```

Query Match	27.5%	Score 33.8;	DB 9;	Length 154195;
Best Local Similarity	56.9%;	Pred. No. 0.54;		
Matches 62;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0;

Qy 68 gactgtaagttcaccatgtaagtgataactatcttcaattattcca 116  
| | | | | | | | | | | | | | | | | | | | | |  
Db 33862 GTGTTTAATTAACAATGATTTGTGTCTTATTTTGAAGCTATTCCA 33814

Db 33862 GTGTTTAATTAAACATGATTGCTGCTTATTTTGAAGCTATCCA 33814

RESULT	14
LOCUS	AP001146/c
DEFINITION	AP001146 166266 bp DNA HTG 30-MAY-2000
ACCESSION	Homo sapiens chromosome 11 clone RP11-727G17 map 11q22, WORKING DRAFT SEQUENCE, 16 unordered pieces.
VERSION	AP001146
KEYWORDS	AP001146.2 GI:8118522
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT
ORGANISM	Homo sapiens DNA, clone:RP11-727G17.
	Homo sapiens

REFERENCE  
AUTHORS

Eukaryota; Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 166266)  
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,

**TITLE** Homo sapiens 166,266 genomic DNA of 11q22  
**JOURNAL** Published only in Database (2000) in press  
**REFERENCE** 2 (bases 1 to 166265)  
**AUTHORS** Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Torokii,Y., Watanabe,H. and Sakaki,Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (08-FEB-2000) Masahira Hattori, The Institute of Physical

COMMENT  
On May 31, 2000 this sequence version replaced gi:6997824.  
Fax: 81-42-778-9924  
URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-42-778-9933,  
Japan (E-mail: [nattori@gsc.riken.go.jp](mailto:nattori@gsc.riken.go.jp),  
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555

----- Genome Center  
Center: RIKEN Genomic Sciences Center(GSC)  
Center code: RIKEN  
Web site: <http://hnp.gsc.riken.go.jp/>  
Contact: [hatori@gsc.riken.go.jp](mailto:hatori@gsc.riken.go.jp)  
----- Project Information  
Center project name: HumDrat11  
Center clone name: Rp11-727617

```

----- Summary Statistics: 100% of reads
Sequencing vector: PCR products: 100% of reads
Chemistry: Dye-terminator ET-amersham: 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 155729 bases at least Q40
Consensus quality: 160730 bases at least Q30
Consensus quality: 162836 bases at least Q20
Insert size: 164766; sum-of-configs
Quality coverage: 5.73x in Q20 bases; sum-of-configs

```

NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

Contig	Length	Contig	Length	Contig	Length
43738	6093	contig of	16336	bp in	length
60154	77413	contig of	17220	bp in	length
77514	94208	contig of	16655	bp in	length
94309	107738	contig of	13400	bp in	length
107889	117065	contig of	9227	bp in	length
117166	124047	contig of	6882	bp in	length
124148	131440	contig of	7293	bp in	length
131514	138828	contig of	7288	bp in	length
138929	144091	contig of	5163	bp in	length
144192	150955	contig of	6754	bp in	length
151056	156185	contig of	5130	bp in	length
156266	158976	contig of	2691	bp in	length
159077	162039	contig of	3023	bp in	length
162200	165096	contig of	2897	bp in	length







```
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: l8126
Center clone name: 304-I11
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158038 bases at least Q40
Consensus quality: 173087 bases at least Q30
Consensus quality: 177952 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 180004; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 9433: contig of 9433 bp in length
* 9434 9533: gap of 100 bp
* 9534 10744: contig of 1211 bp in length
* 10745 10844: gap of 100 bp in length
* 10845 12149: contig of 1305 bp in length
* 12150 12249: gap of 100 bp
* 12250 13273: contig of 1024 bp in length
* 13274 13373: gap of 100 bp
* 13374 14552: contig of 1179 bp in length
* 14553 14652: gap of 100 bp
* 14653 15679: contig of 1027 bp in length
* 15680 15779: gap of 100 bp
* 15780 16942: contig of 1163 bp in length
* 16943 17042: gap of 100 bp
* 17043 18595: contig of 1553 bp in length
* 18596 18695: gap of 100 bp
* 18696 19867: contig of 1172 bp in length
* 19868 19967: gap of 100 bp
* 19968 20972: contig of 1005 bp in length
* 20973 21072: gap of 100 bp
* 21073 22260: contig of 1188 bp in length
* 22261 22360: gap of 100 bp
* 22361 23548: contig of 1188 bp in length
* 23549 23648: gap of 100 bp
* 23649 24920: contig of 1272 bp in length
* 24921 25020: gap of 100 bp
* 25021 26233: contig of 1213 bp in length
* 26234 26333: gap of 100 bp
* 26334 27853: contig of 1520 bp in length
* 27854 27953: gap of 100 bp
* 27954 29293: contig of 1340 bp in length
* 29294 29393: gap of 100 bp
* 29394 53904: contig of 24511 bp in length
* 53905 54004: gap of 100 bp
* 54005 55149: contig of 1145 bp in length
* 55150 55249: gap of 100 bp
* 55250 56417: contig of 1168 bp in length
* 56418 56517: gap of 100 bp
* 56518 57905: contig of 1388 bp in length
* 57906 58005: gap of 100 bp
* 58006 59297: contig of 1292 bp in length
* 59298 59397: gap of 100 bp
* 59398 60783: contig of 1386 bp in length
* 60784 60883: gap of 100 bp
* 60884 62491: contig of 1608 bp in length
* 62492 62591: gap of 100 bp
* 62592 64008: contig of 1417 bp in length
* 64009 64108: gap of 100 bp
* 64109 65352: contig of 1244 bp in length
```

```
* 65353 65452: gap of 100 bp
* 65453 66782: contig of 1330 bp in length
* 66783 66882: gap of 100 bp
* 66883 68395: contig of 1513 bp in length
* 68396 68495: gap of 100 bp
* 68496 70756: contig of 2261 bp in length
* 70757 70856: gap of 100 bp
* 70857 72429: contig of 1573 bp in length
* 72430 72529: gap of 100 bp
* 72530 73566: contig of 1037 bp in length
* 73567 73666: gap of 100 bp
* 73667 74917: contig of 1251 bp in length
* 74918 75017: gap of 100 bp
* 75018 77012: contig of 1995 bp in length
* 77013 77112: gap of 100 bp
* 77113 79544: contig of 2432 bp in length
* 79545 79644: gap of 100 bp
* 79645 81091: contig of 1447 bp in length
* 81092 81191: gap of 100 bp
* 81192 83043: contig of 1852 bp in length
* 83044 83143: gap of 100 bp
* 83144 85037: contig of 1894 bp in length
* 85038 85137: gap of 100 bp
* 85138 87330: contig of 2193 bp in length
* 87331 87430: gap of 100 bp
* 87431 89273: contig of 1843 bp in length
* 89274 89373: gap of 100 bp
* 89374 90847: contig of 1474 bp in length
* 90848 90947: gap of 100 bp
* 90948 93191: contig of 2244 bp in length
* 93192 93291: gap of 100 bp
* 93292 95994: contig of 2703 bp in length
* 95995 96094: gap of 100 bp
* 96095 99821: contig of 3727 bp in length
* 99822 99921: gap of 100 bp
* 99922 103654: contig of 3733 bp in length
* 103655 103754: gap of 100 bp
* 103755 107159: contig of 3405 bp in length
* 107160 107259: gap of 100 bp
* 107260 110785: contig of 3526 bp in length
* 110786 110885: gap of 100 bp
* 110886 115380: contig of 4495 bp in length
* 115381 115480: gap of 100 bp
* 115481 119514: contig of 4034 bp in length
* 119515 119614: gap of 100 bp
* 119615 123779: contig of 4165 bp in length
* 123780 123879: gap of 100 bp
* 123880 133095: contig of 9216 bp in length
* 133096 133195: gap of 100 bp
* 133196 139671: contig of 6476 bp in length
* 139672 139771: gap of 100 bp
* 139772 147743: contig of 7972 bp in length
* 147744 147843: gap of 100 bp
* 147844 156603: contig of 8760 bp in length
* 156604 156703: gap of 100 bp
* 156704 169229: contig of 12526 bp in length
* 169230 169329: gap of 100 bp
* 169330 185113: contig of 15764 bp in length
* 185114 185213: gap of 100 bp
* 185214 185404: contig of 191 bp in length.
*
* location/Qualifiers
* 1. 185404
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="1"
* /map="1"
* /clone_lib="RP11-304111"
* /clone_1lb="RP11 Human Male BAC"
* 1. 9433
* /note="assembly_fragment"
* clone_end:sp6
* vector_side:left"
* 9534 .10744
```

misc\_feature

misc\_feature

## FEATURES

source







QY 11 aatgtgagatgcctaccacccctgacgaatgacgagggctgacac 70  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 65951 ATATGTCAGATATTCTGACCTTAATTAATGAGGGAATGCTGTAAGAC 66010  
QY 71 tggtaagttcacatgtagtgatctatctcaattatccacaca 120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 66011 TGGTTCCAGAACATTTGAGGAGCTCATTTATAGCTTTGTTAACAAA 66060

RESULT 17  
GGGP23 799 bp mRNA VRT 24-MAR-1992  
LOCUS Chicken mRNA for microsomal signal peptidase.  
ACCESSION X60795  
VERSION X60795.1 GI:63421  
KEYWORDS microsomal signal peptidase; signal peptidase.  
SOURCE Chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 799)  
AUTHORS Lively,M.O.  
JOURNAL Direct Submission  
TITLE Submitted (05-JUL-1991) M.O. Lively, Bowman Gray School of Medicine  
of, Wake Forest University, dept of Biochemistry, Medical Center  
Blvd, Winston-Salem NC 27157, USA  
2 (bases 1 to 799)  
REFERENCE Newsome,A.L., McLean,J.W. and Lively,M.O.  
AUTHORS Molecular cloning of a cDNA encoding the glycoprotein of hen  
oviduct microsomal signal peptidase  
JOURNAL Biochem. J. 282 (Pt 2), 447-452 (1992)  
MEDLINE 92189580  
FEATURES  
source location/Qualifiers  
1. 799  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/dev\_stage="adult"  
/tissue\_type="oviduct"  
/cell\_type="tubular gland cell"  
/clone\_lib="lambda gt11"  
/clone="56Eco"  
1. 799  
/note="microsomal signal peptidase"  
/evidence="experimental"  
69. 611  
/codon\_start=1  
/evidence="experimental"  
/product="microsomal signal peptidase"  
/protein\_id="CAA4308.1"  
/db\_xref="GI:63422"  
/db\_xref="SWISS-PROT:P28667"  
/translation="NMTVLSRANSLSFSLISVMAALTFPCFTTAKRSKRSVPISAVS  
RVLIRNVEFTGPRERSDLAFEDITADLSEFDNNVKQLFLYLSAEYSTKNALNO  
VVLMDKILIRGNRPRLFLDKMSKRYFFPDGSLKNNRVITLISNNVYPNGLPLV  
TSGHMSVPPFPYETTKSY"  
765. 770  
polya\_signal /evidence="experimental"  
BASE COUNT 230 a 182 c 174 g 213 t  
ORIGIN

Query Match 25.5%; Score 31.4; DB 5; Length 799;  
Best Local Similarity 54.9%; Pred. No. 2.8;  
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 11 aatgtgagatgcctaccacccctgacgaatgacgagggctgacac 70  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 246 AGAAGTCATCTGGCGTTCGTCACGTTTGACATTAATGCGAGAGTATATTGGAC 305  
QY 71 tggtaagttcacatgtagtgatctatctcaattatccacacaca 123  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 306 TGAATGTAAACAATTTGTTCTAATTTGTCGAGAAATATTCACAAAAA 358  
RESULT 18  
AE002311  
LOCUS AE002311 23366 bp DNA BCT 26-MAY-2000  
DEFINITION Chlamydia muridarum, section 42 of 85 of the complete genome.  
ACCESSION AE002311 AE002160  
VERSION AE002311.1 GI:7190477  
KEYWORDS  
SOURCE Chlamydia muridarum.  
ORGANISM Chlamydia muridarum.  
REFERENCE Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
AUTHORS 1. (bases 1 to 23366)  
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Unayan,L.A., Uterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
JOURNAL 20150255  
MEDLINE 10684935  
PUBMED  
REFERENCE 2 (bases 1 to 23366)  
AUTHORS Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Unayan,L.A., Uterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Direct Submission  
TITLE Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source location/Qualifiers  
1. 23366  
/organism="Chlamydia muridarum"  
/db\_xref="taxon:83560"  
/note="synonym: Chlamydia trachomatis Mohn"  
complement(136. 1230)  
/gene="TC0435"  
complement(136. 1230)  
/gene="TC0435"  
/note="similar to GP:2444077; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="phospholipase D family protein"  
/protein\_id="AF39289.1"  
/db\_xref="GI:7190478"  
/translation="MVSPLITTSPTLTLTLQARLFTLTSSSSONOKTCARKIOELT  
RPSLLALVRSPLSSEYCFPSKCGYRNLADLIEKVIISATSIYLKNNVPSSKSLA  
OALILALOGVPSYIHPKHTLKALDLSQTNVLTYSRASSDKOTLVLDHQQVVG  
ARNFYTTSLHREANLMKRLISLQNLAMLEKNQGEVCLQGYCYCPISKNKGGSEK  
ELIKEOKATSIQIGTSCITQSQIILADENAVTRVIVITIDSNNKYTLERLSLK  
CYRINRVCGRSSSNOCVCIHDTITAIIGSSSRPFRROANQEAVALIISPLETCQKO  
DICTMNRSLDSCSTATTYEAEADSSSE"  
complement(1247. 2431)  
/gene="TC0436"  
/gene="TC0436"  
complement(1247. 2431)  
/note="similar to GB:AE001363; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="phospholipase D family protein"  
/protein\_id="AF39290.1"  
/db\_xref="GI:7190479"  
/translation="MDPIRGPHQKSSKRSKIRSFSPMRINETHLVLOERPLSMGGA  
VSSLSFSPWTPARTLTITSPYQOKLFTFKYOKLTPPSALATIPYNNSSSSICPS  
KKGDAIVNITFAILISAKTSIFLKIFISSKYKXIKRPLSSALISQGVSM/LHHHSK  
TVSDSKTTTIELIIOENSHALHNKTLIIDKQOIIITLNGNFSASLCKDVLNMIIN  
NMHLASLMEKNAGRACIKOKVRYLPVNRKGIIVEIQRASSIOLGMYIITLNEAI  
KALDEAASORSVLTIIIDISITKQOTLFTLKALNSKIRVRAGTALASCIHCKYCIIDHK



gene  
CDS

TVIGSANSWNSGRLNTNKEDLLIISPLTEGOKODICTWVHFLICENSTVLAESTKTPKI  
LD"  
3255..13022  
/gene="TC0437"  
3255..13022  
/gene="TC0437"  
/note="similar to GP:6013469; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="adherence factor"  
/protein\_id="AAF39291.1"  
/db\_xref="GI:7190480"

gene  
CDS

EENONHTEDVAVRIGCAMLOOQFYNNMTGRNESFTYPAHSGMKTSLIYNQAQVAH  
IFSLNTYPAKIIHHNGTKOMESKOPVYSSEKSNKRDVSEIAYORVNEVETKS  
SLAKTASIRRRKPREVEERKBDLVKDPADKIYDRLADPTIASRLKTEBOIOKOKISL  
NLKATORYLINDSEBNSRGGODLYKOSFLLNLIENAMESSEBKVNMSYIKT  
EPNSHRXKIDSHFHGTWAGAPBECTDYIKTFLLDTTEFDFLEWDRQAFANFSS  
ILKRIAFDSSLKELKSLTSQETQDVKYDELOAKYQTSGRREERLELEKRLKLEDR  
YKHLQSELRITNEDALFLKNVLSODGFENYCLKIGTIDQTRAYLODELKLEBE  
IOQYKDLIKANEOKIKDIKGVYNTDLGKERYIKNIRLDSKQDRTNYNEMMLR  
WVYPAASDOIWMYMLKELGGIYTDLDMPROYSPVLOMINETIGDGFPELPTIRAVS  
DGYRLANGESGITTDNIADODISKLSRSDPTLEKILLTDENOKSGSTSESLFO  
RMADSVRDEPMLIQRHKSTGAVNRGLNGIMAHKKSAMDAVIRKORARIELKS  
LRETVLSEEFRTLDLKLHNLKALIGHLVEDFLGSLFSEFRODTIAGALSTLGI  
TGPDLYDEMERTYRSLGPIGKDYLEGKRLGKAFGLSYNKLVKEGKOTFDMLHEVTY  
GANVTPADESTWCGIKORLAELFEDSEKFRPTPKGISRTKIDKFTFLMEATY  
KRALSPEDLDEFTNGLIEAKNFDLIKETFDYDALYVLRNOVCDAASAVAFILOLA  
ELRGTSPIANLONVPPDLNRNTASSLEKTKLYLOSHSQERYVIMSAISDRVMPL  
RDMLATARMPLVKESTSEKERTPLTTEIKILNMAELQSKDNIGILSEVYDMLD  
VTSIMESKILKRVSAIEDSTVAGTFYLRLEETISEWRIPEPDLKRVLRFMGMD  
ANQHPDEGRGWYSQLYDESFKRVYPAVKIQLAKRYLMEQVHLEIDYLTKN  
PLFLRHEGEAFSDLEITRYMLAEVIGSIFSEGNIPLPSARLVNIKITYYGDY  
HDMODVLAIKYDMLASGAGADLTNERFAIPESLRNLEGLTGDLPLPPVAVSG  
GMSGFENGVSVDLMISIAQFENGASYSMOHTYALYEIHRHILGSLTDLKKE  
LESKAGCGVHEERPDLSLKASSEKOVSLTEIKHSINOVHIAVSHIMNTALPY  
GKITERBEDFRGLGTTTETYSANRSYTRIGIYSKDLPSLPOEYPTIONVEOKAY  
LLSWPEFYNRANAMASDLATHYGAELIENHPOSFLYEVGRCMQJSLIYMSLEDGGY  
RTLOGNDLTVALYQOKERDHLPLSKDOSLSDLSLIMLOYGNVPLENRPNH  
AKMVLQITSTFEKSLTSLITTPTHSLTNFMGSPFRVDPNGHVDPLAARF  
FIEDMVOYSPHIKORYGSESLTSYREOLOVYFLADPEGOSLSTFDGLSHOAT  
LEKREBPVATIERIOTWGSLLTYSRGGGTIYSHHRRIDEPTSEIDPMKINGDVLPEIT  
KHVIDSEVSLIIVLETRGLEBGRKHSRSIYVTPDEASLFOFLKNKRAKTEML  
VSLISEVYDRLKSAQDGEDVEVSRSFSDNDEITVEVKKDPLILKLSIRGGJFDS  
FKQPSANENLGAQVMDLELGMYSIVQYMRLEBEGKGPLANAMVDMOKETAE  
MTGAVIOGLSKFTITQBGVGFLETAIASOLKASSRVGSVYKSKPAVLELP  
VLETAIGAMNLYDSVSLKEAOTHSDMAARVDIANSITLIGITIASVAPSLMAYG  
PLAIGMCAASAIARNVAKREYDQMLYKFKITRGSGHILFADPDRLHLDSSGYVL  
GNFLDLDRONPPLIKGDRSYNSNRITGKAGNTMOQIDRIGYIRISPTSLARGHA  
NSFWPELPPTIPKDYRTVILGYTTYEAVTEVYLSNRYVREAVMSSSKYRPPPL  
TAKSKLSIIIGDSDPLVYIPVRLIEKDEKENIDHNASKDKINKKGGGLTVAIG  
AGFYNLGTLPATNITSPRAIPAPGVFNLSIVODVPLIRPNCTEKKFLIRDTGF  
NTILGSEGNQTLTGNDTKFYSSPGGTYIHKHVSVEYHLPOLHPIHILISKSSA  
HOLYTEINAQOVKPYVLSLSTSSMELISKOEIRHSSDTEKDELGWKKEREVRLGDC  
ITATPIDMNAQTRVLWGVKTCDAKQKOPHETSTYPEDILDLVTKRSMAYELV  
FVLENGTAKFTNKRHIYSPKPYSELAVTOSRDKAVQNGTGSYFATYRSQK  
TOQTELELVDDVDYAOQIDVSSILPTLLAKSGDNQIDFEFSSHHLMATIRYSOK  
IPEKTRVKNINNETILRGLMNAKLAMASNPXYMLLYNRYNIPERVADYSLNNTV  
LMPNTRKNGEHVAIEMOEGVDVKGIFLOSCHVGLYEEDIMAKYDPAKITHFAV  
TYPAKTKYLEFTDEESLANPEDVNLPSKIGSNTLOASQKPLTOFAOLMAKSTDEIS  
VSDTSKLENFARYLMEAEETILEYRQLLYAONLILNHRDVLVKLEYROSGMAVL  
ITFRNFTIESMEGSIAPTLEREAKPIILASDGMPLIRREHDLFLEDCIKMLIAVL  
QOEFMLOYIVAMOGKALGSLTPPYKDMMLAVLYTVVASMEBRASMLKAIIEY  
KLPLETTLTASYLYLXPVSGDLVYTCIMLRPONOAVYRFKFGYKOKKAPASIALISS  
HSVGESRTAKVLOPFGOELFHLKSAPNKSAVYLODLSKLSISMLLEYGGDQVIRY  
PLSYOQFLSDDLMVADLKERVMANEGEFLISOMHSHLTIEMESOMKRDVAPKIN  
PNTVINAITYYRVVVPWIRNRLTKGSLIRIPARSIPALITTTODNVEKTIYAPESL  
YFSEFLGDAFLKPRTHNRVGDMLLEKEVVVTVYKVKDETDYNNKRVYIVELTEKEE  
RVLKANKNVTIRIPKSSSYTKK"

/gene="TC0438"  
/note="similar to GP:6013469; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="adherence factor"  
/protein\_id="AAF39292.1"  
/db\_xref="GI:7190481"

FTNHTEDVAVRIGCAMLOOQFYNNMTGRNESFTYPAHSGMKTSLIYNQAQVAH  
IFSLNTYPAKIIHHNGTKOMESKOPVYSSEKSNKRDVSEIAYORVNEVETKS  
SLAKTASIRRRKPREVEERKBDLVKDPADKIYDRLADPTIASRLKTEBOIOKOKISL  
NLKATORYLINDSEBNSRGGODLYKOSFLLNLIENAMESSEBKVNMSYIKT  
EPNSHRXKIDSHFHGTWAGAPBECTDYIKTFLLDTTEFDFLEWDRQAFANFSS  
ILKRIAFDSSLKELKSLTSQETQDVKYDELOAKYQTSGRREERLELEKRLKLEDR  
YKHLQSELRITNEDALFLKNVLSODGFENYCLKIGTIDQTRAYLODELKLEBE  
IOQYKDLIKANEOKIKDIKGVYNTDLGKERYIKNIRLDSKQDRTNYNEMMLR  
WVYPAASDOIWMYMLKELGGIYTDLDMPROYSPVLOMINETIGDGFPELPTIRAVS  
DGYRLANGESGITTDNIADODISKLSRSDPTLEKILLTDENOKSGSTSESLFO  
RMADSVRDEPMLIQRHKSTGAVNRGLNGIMAHKKSAMDAVIRKORARIELKS  
LRETVLSEEFRTLDLKLHNLKALIGHLVEDFLGSLFSEFRODTIAGALSTLGI  
TGPDLYDEMERTYRSLGPIGKDYLEGKRLGKAFGLSYNKLVKEGKOTFDMLHEVTY  
GANVTPADESTWCGIKORLAELFEDSEKFRPTPKGISRTKIDKFTFLMEATY  
KRALSPEDLDEFTNGLIEAKNFDLIKETFDYDALYVLRNOVCDAASAVAFILOLA  
ELRGTSPIANLONVPPDLNRNTASSLEKTKLYLOSHSQERYVIMSAISDRVMPL  
RDMLATARMPLVKESTSEKERTPLTTEIKILNMAELQSKDNIGILSEVYDMLD  
VTSIMESKILKRVSAIEDSTVAGTFYLRLEETISEWRIPEPDLKRVLRFMGMD  
ANQHPDEGRGWYSQLYDESFKRVYPAVKIQLAKRYLMEQVHLEIDYLTKN  
PLFLRHEGEAFSDLEITRYMLAEVIGSIFSEGNIPLPSARLVNIKITYYGDY  
HDMODVLAIKYDMLASGAGADLTNERFAIPESLRNLEGLTGDLPLPPVAVSG  
GMSGFENGVSVDLMISIAQFENGASYSMOHTYALYEIHRHILGSLTDLKKE  
LESKAGCGVHEERPDLSLKASSEKOVSLTEIKHSINOVHIAVSHIMNTALPY  
GKITERBEDFRGLGTTTETYSANRSYTRIGIYSKDLPSLPOEYPTIONVEOKAY  
LLSWPEFYNRANAMASDLATHYGAELIENHPOSFLYEVGRCMQJSLIYMSLEDGGY  
RTLOGNDLTVALYQOKERDHLPLSKDOSLSDLSLIMLOYGNVPLENRPNH  
AKMVLQITSTFEKSLTSLITTPTHSLTNFMGSPFRVDPNGHVDPLAARF  
FIEDMVOYSPHIKORYGSESLTSYREOLOVYFLADPEGOSLSTFDGLSHOAT  
LEKREBPVATIERIOTWGSLLTYSRGGGTIYSHHRRIDEPTSEIDPMKINGDVLPEIT  
KHVIDSEVSLIIVLETRGLEBGRKHSRSIYVTPDEASLFOFLKNKRAKTEML  
VSLISEVYDRLKSAQDGEDVEVSRSFSDNDEITVEVKKDPLILKLSIRGGJFDS  
FKQPSANENLGAQVMDLELGMYSIVQYMRLEBEGKGPLANAMVDMOKETAE  
MTGAVIOGLSKFTITQBGVGFLETAIASOLKASSRVGSVYKSKPAVLELP  
VLETAIGAMNLYDSVSLKEAOTHSDMAARVDIANSITLIGITIASVAPSLMAYG  
PLAIGMCAASAIARNVAKREYDQMLYKFKITRGSGHILFADPDRLHLDSSGYVL  
GNFLDLDRONPPLIKGDRSYNSNRITGKAGNTMOQIDRIGYIRISPTSLARGHA  
NSFWPELPPTIPKDYRTVILGYTTYEAVTEVYLSNRYVREAVMSSSKYRPPPL  
TAKSKLSIIIGDSDPLVYIPVRLIEKDEKENIDHNASKDKINKKGGGLTVAIG  
AGFYNLGTLPATNITSPRAIPAPGVFNLSIVODVPLIRPNCTEKKFLIRDTGF  
NTILGSEGNQTLTGNDTKFYSSPGGTYIHKHVSVEYHLPOLHPIHILISKSSA  
HOLYTEINAQOVKPYVLSLSTSSMELISKOEIRHSSDTEKDELGWKKEREVRLGDC  
ITATPIDMNAQTRVLWGVKTCDAKQKOPHETSTYPEDILDLVTKRSMAYELV  
FVLENGTAKFTNKRHIYSPKPYSELAVTOSRDKAVQNGTGSYFATYRSQK  
TOQTELELVDDVDYAOQIDVSSILPTLLAKSGDNQIDFEFSSHHLMATIRYSOK  
IPEKTRVKNINNETILRGLMNAKLAMASNPXYMLLYNRYNIPERVADYSLNNTV  
LMPNTRKNGEHVAIEMOEGVDVKGIFLOSCHVGLYEEDIMAKYDPAKITHFAV  
TYPAKTKYLEFTDEESLANPEDVNLPSKIGSNTLOASQKPLTOFAOLMAKSTDEIS  
VSDTSKLENFARYLMEAEETILEYRQLLYAONLILNHRDVLVKLEYROSGMAVL  
ITFRNFTIESMEGSIAPTLEREAKPIILASDGMPLIRREHDLFLEDCIKMLIAVL  
QOEFMLOYIVAMOGKALGSLTPPYKDMMLAVLYTVVASMEBRASMLKAIIEY  
KLPLETTLTASYLYLXPVSGDLVYTCIMLRPONOAVYRFKFGYKOKKAPASIALISS  
HSVGESRTAKVLOPFGOELFHLKSAPNKSAVYLODLSKLSISMLLEYGGDQVIRY  
PLSYOQFLSDDLMVADLKERVMANEGEFLISOMHSHLTIEMESOMKRDVAPKIN  
PNTVINAITYYRVVVPWIRNRLTKGSLIRIPARSIPALITTTODNVEKTIYAPESL  
YFSEFLGDAFLKPRTHNRVGDMLLEKEVVVTVYKVKDETDYNNKRVYIVELTEKEE  
RVLKANKNVTIRIPKSSSYTKK"

Query Match		25.4%	Score 31.2;	DB 1;	Length 23366;
Best Local Similarity		55.6%	Pred. No. 4;		
Matches 60;		Conservative 0;	Mismatches 48;	Indels 0;	Gaps 0;
Qy	10	taaaagtgaagcttcacacccgagcttcacacccgagcgaacgagggctgcaciga	69		
Db	9518	TAAGGAGATCGCCTCTTAATAATTCGAGATAGGATAGGATAGGCTGATGACAGA	9577		
Qy	70	ctggtgaagttcacatgtaagtgatacttacttcaatcttaccac	117		
Db	9578	CTGGCAAAATTCGAGATAGATAGATTAGGATTATTCGATTCGACG	9625		



RESULT 19  
CEK10D6 30494 bp DNA INV 25-OCT-2000  
LOCUS Caenorhabditis elegans cosmid K10D6, complete sequence.  
DEFINITION  
ACCESSION Z74040  
VERSION Z74040.1 GI:1403245  
KEYWORDS HTG: GABA receptor.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 30494)  
none.  
REFERENCE  
AUTHORS none.  
TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
JOURNAL Science 282 (5396), 2012-2018 (1998)  
MEDLINE 99069613  
REMARK The C.elegans Sequencing Consortium.  
Erratum:[published errata appear in Science 1999 Jan  
1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep  
3;285(5433):1493]]  
2 (bases 1 to 30494)  
White,S.  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (14-JUN-1996) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RO, England and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or twenematode.wustl.edu  
Coding sequences below are predicted from computer analysis, using  
predictions from GeneFinder (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is not the entire insert of clone K10D6.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone K10D6 is at 1 in this sequence. The true  
left end of clone T19B10 is at 30388 in this sequence. The start of  
this sequence (1..108) overlaps with the end of sequence AL021173.  
The end of this sequence (30388..30494) overlaps with the start of  
sequence Z74043.  
For a graphical representation of this sequence and its analysis  
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=K10D6)  
name=K10D6.  
FEATURES  
source  
Location/Qualifiers  
1..30494  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="V"  
/clone="K10D6"  
join(1925..2071,3486..3705,3814..3893,3953..4045,  
4098..4174,4480..4528,4580..4905,4954..5404,5462..5539)  
/gene="K10D6.1"  
join(1925..2071,3486..3705,3814..3893,3953..4045,  
4098..4174,4480..4528,4580..4905,4954..5404,5462..5539)  
/note="predicted using GeneFinder  
Similarity to Drosophila GABA receptor beta subunit  
(SW:GAB DROME), contains similarity to Pfam domain:  
PF00065 (Neurotransmitter-gated ion-channel), Score=148.6,  
E-value=5.7e-43, N=2"  
/codon\_start=1  
/protein\_id="CA98509.1"  
/db\_xref="GI:3878477"  
/translation="MGIFLINISIFPIVPSFIFQLSMTFLDPNKGSGFIYHDY  
MSSQHNKRTNSFFIFLILSIDINCEITIDSDIVSSDDPTTQKRNCTR  
OPAIIDLKNGTGYNFRIPENGVQVVEFWQATISINETINDEMDIYINEMILD  
PAIKFDHLNPKONLSVSHOVLEQIWTPNSCFVNSKFAEIHDSPFKNVFLMIYPNCTV

WNYRVNKGPCDLSLEFLPDIOECHLIYESFNYNQGYRMKRNKESAPSVSTNKI  
RLDPFELIKIESTRIAPYAGMDEIHLVFEPRITWFMQAYLPTYLITFISMS  
FSLGTAKMPARTMIGVALLAMIFOGNIRNRPVSYVAIDVMIVSMFIFLSIL  
ELAIIVGYTKTNEEGSKKCPHKIKLLDNFESPGICRKYERKFLPERSARNGGIIR  
FOYQDFWNMSPEKIDRVSAIMPACFAFINIYVMSYYIKKLEKAAEMNDRL"  
complement(join(6124..6262,6777..6830,6977..7187,  
7236..7461,7969..8202))  
/gene="K10D6.2b"  
7236..7461,7969..8202))  
complement(join(6124..6262,6777..6830,6977..7187,  
7236..7461,7969..8202))  
/gene="K10D6.2b"  
/note="cDNA EST yk28494.3 comes from this gene  
cDNA EST yk28494.5 comes from this gene  
cDNA EST yk525b5.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CA98513.1"  
/db\_xref="GI:3878481"  
/db\_xref="SPRMBL:045669"  
/translation="MGSTVQKLLALISYLIVGICITSGCSPAMQVVDIREFRA  
EHOGHLMDCIRAEKHVSVGDEYDEPLHGMKFPNSAELVIONTLNIDEGAGE  
SEHHRFAMWKALIFLITSEFLAFISICGVCAPCPSTAPFISLFIAMCSILA  
DGIFFLAANRVDRFVQGWGTYEQRIGAFYIHLMGTLCMIGAFVCTLTITTKPYTG  
DEEDGSKENLFTWQOQTOPVTLVSLVCEVFCWASLSFQKIRSSSAATVSSQOQ  
TSITQYRETSA"  
complement(join(6150..6262,6410..6594,6777..6830,  
6977..7187,7236..7461,7969..8202))  
/gene="K10D6.2a"  
complement(join(6150..6262,6410..6594,6777..6830,  
6977..7187,7236..7461,7969..8202))  
/gene="K10D6.2a"  
/note="cDNA EST yk255d4.3 comes from this gene  
cDNA EST yk255d4.5 comes from this gene  
cDNA EST yk526b2.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CA98514.2"  
/db\_xref="GI:6425186"  
/db\_xref="SPRMBL:045670"  
/translation="MGSTVQKLLALISYLIVGICITSGCSPAMQVVDIREFRA  
EHOGHLMDCIRAEKHVSVGDEYDEPLHGMKFPNSAELVIONTLNIDEGAGE  
SEHHRFAMWKALIFLITSEFLAFISICGVCAPCPSTAPFISLFIAMCSILA  
DGIFFLAANRVDRFVQGWGTYEQRIGAFYIHLMGTLCMIGAFVCTLTITTKPYTG  
DEEDGSKENLFTWQOQTOPVTLVSLVCEVFCWASLSFQKIRSSSAATVSSQOQ  
HTEQAPV"  
complement(join(6262,6406..6594,6977..7187,7236..7461,  
7969..8202))  
/gene="K10D6.2c"  
complement(join(6262,6406..6594,6977..7187,7236..7461,  
7969..8202))  
/gene="K10D6.2c"  
/codon\_start=1  
/protein\_id="CAB54276.2"  
/db\_xref="GI:6425185"  
/translation="MGSTVQKLLALISYLIVGICITSGCSPAMQVVDIREFRA  
EHOGHLMDCIRAEKHVSVGDEYDEPLHGMKFPNSAELVIONTLNIDEGAGE  
SEHHRFAMWKALIFLITSEFLAFISICGVCAPCPSTAPFISLFIAMCSILA  
DGIFFLAANRVDRFVQGWGTYEQRIGAFYIHLMGTLCMIGAFVCTLTITTKPYTG  
DEEDGSKENLFTWQOQTOPVTLVSLVCEVFCWASLSFQKIRSSSAATVSSQOQ  
HTEQAPV"  
complement(join(13294..13356,13407..13595)  
join(13294..13356,13407..13595)  
/gene="K10D6.3"  
/note="predicted using GeneFinder"  
/codon\_start=1  
/protein\_id="CA98510.1"  
/db\_xref="GI:3878478"  
/db\_xref="SPRMBL:021421"  
/translation="MSTQSPHSSNDLSICTARIEVSVPRKPKNSTTPVARY  
EMKSKESDVDAKLVKSSLDCKLEKRAPTCCKAA".27301,27401..27463,  
27716..27902,27954..28159,28296..28568)  
/gene="K10D6.4"







Query Match	Best Local Similarity	Score	DB	Length
Matches 60; Conservative %	0; Mismatches	50; Indels	0; Gaps	0;
QY 12	aatgtagagatgccttaacaacccatgaataatgacgaatgcgaagagctgcactgact	71		
Db 72816	AATGAGAAATTTCCACTCGGCAAGATGACCATGTGATATTACAGAAGCCTTGTGCTACT	72757		
QY 72	ggtatgattccatcgaatgaagtgatcactatcttaattcattccacaaca	121		
Db 72756	TGTAAAGGAGCAAAATGTGAGGTCATCTCTTTCTTTCTTTACAAACAA	72707		
RESULT 21	AC027346/c			
LOCUS	AC027346	68575 bp	DNA	HTG 25-APR-2001
DEFINITION	Homo sapiens chromosome 16 clone CTD-2519D12, WORKING DRAFT			
ACCESSION	AC027346			
VERSION	AC027346.4			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.			
TITLE	Sequencing of Human Chromosome 16			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 68575)			
AUTHORS	DOE Joint Genome Institute.			
TITLE	Sequencing of Human Chromosome 16			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
AUTHORS	On Apr 25, 2001 this sequence version replaced g1:8576174.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint			
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			



-----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: <http://www.jgi.doe.gov>  
 -----  
 Project Information  
 Center Project Name: 797361  
 Center clone name: CITB-EL\_2519D12  
 -----  
 Summary Statistics  
 Consensus quality: 36726 bases at least Q40  
 Consensus quality: 43164 bases at least Q30  
 Consensus quality: 46452 bases at least Q20  
 Estimated insert size: 53000; agarose-1p estimation  
 Estimated insert size: 65075; sum-of-coverage estimation  
 Quality coverage: 8.56 in Q20 bases; agarose-1p estimation  
 Quality coverage: 6.98 in Q20 bases; sum-of-coverage estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 36 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1031: contig of 1031 bp in length  
 \* 1032 1131: gap of unknown length  
 \* 1132 2253: contig of 1122 bp in length  
 \* 2254 2353: gap of unknown length  
 \* 2354 3402: contig of 1049 bp in length  
 \* 3403 3502: gap of unknown length  
 \* 3503 4718: contig of 1216 bp in length  
 \* 4719 4818: gap of unknown length  
 \* 4819 5849: contig of 1031 bp in length  
 \* 5850 5949: gap of unknown length  
 \* 5950 6964: contig of 1015 bp in length  
 \* 6965 7064: gap of unknown length  
 \* 7065 8110: contig of 1046 bp in length  
 \* 8111 9258: contig of 1048 bp in length  
 \* 9259 9358: gap of unknown length  
 \* 9359 10395: contig of 1037 bp in length  
 \* 10396 10495: gap of unknown length  
 \* 10496 11630: contig of 1135 bp in length  
 \* 11631 11730: gap of unknown length  
 \* 11731 12758: contig of 1028 bp in length  
 \* 12759 12858: gap of unknown length  
 \* 12859 13960: contig of 1102 bp in length  
 \* 13961 14060: gap of unknown length  
 \* 14061 15269: contig of 1209 bp in length  
 \* 15270 15369: gap of unknown length  
 \* 15370 16529: contig of 1160 bp in length  
 \* 16530 16629: gap of unknown length  
 \* 16630 17838: contig of 1209 bp in length  
 \* 17839 17938: gap of unknown length  
 \* 17939 19216: contig of 1278 bp in length  
 \* 19217 19316: gap of unknown length  
 \* 19317 20590: contig of 1274 bp in length  
 \* 20591 20690: gap of unknown length  
 \* 20691 21810: contig of 1120 bp in length  
 \* 21811 21910: gap of unknown length  
 \* 21911 23108: contig of 1198 bp in length  
 \* 23109 23208: gap of unknown length  
 \* 23209 24544: contig of 1336 bp in length  
 \* 24545 24644: gap of unknown length  
 \* 24645 25781: contig of 1137 bp in length  
 \* 25782 25881: gap of unknown length  
 \* 25882 27081: contig of 1200 bp in length  
 \* 27082 27181: gap of unknown length  
 \* 27182 28208: contig of 1027 bp in length  
 \* 28209 28308: gap of unknown length  
 \* 28309 29677: contig of 1369 bp in length  
 \* 29678 29777: gap of unknown length

29778 30803: contig of 1026 bp in length  
 \* 30804 30903: gap of unknown length  
 \* 30904 32229: contig of 1326 bp in length  
 \* 32230 32329: gap of unknown length  
 \* 32330 33559: contig of 1230 bp in length  
 \* 33560 33659: gap of unknown length  
 \* 33660 33550: contig of 1691 bp in length  
 \* 33551 35450: gap of unknown length  
 \* 35451 36967: contig of 1517 bp in length  
 \* 36968 37067: gap of unknown length  
 \* 37068 38732: contig of 1665 bp in length  
 \* 38733 38832: gap of unknown length  
 \* 38833 39897: contig of 1065 bp in length  
 \* 39898 39997: gap of unknown length  
 \* 39998 41689: contig of 1692 bp in length  
 \* 41690 41789: gap of unknown length  
 \* 41790 43098: contig of 1309 bp in length  
 \* 43099 43198: gap of unknown length  
 \* 43199 44501: contig of 1303 bp in length  
 \* 44502 44601: gap of unknown length  
 \* 44602 45941: contig of 1340 bp in length  
 \* 45942 46041: gap of unknown length  
 \* 46042 68575: contig of 22534 bp in length.  
 FEATURES  
 source location/Qualifiers  
 1..68575  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="C7P-2519D12"  
 /clone\_11b="Caltech human BAC library D"  
 BASE COUNT 18940 a 14129 c 13231 g 18454 t 3821 others  
 ORIGIN

Query Match 24.2%: Score 29.8; DB 2; Length 68575;  
 Best Local Similarity 54.0%; Freq. No. 13;  
 Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 2 agtattgtaaatgtagatgagcttaccacccctgtagcctaattgacgcaatgcgaagc 61  
 Db 60723 AGTACTTGAAACTGTATTATTCACCCCTCTGTCACCTTGCTGTCATGCGACAGC 60664  
 Oy 62 tgcactgactgtagattcacatgtaagtgatactattatcttaattatc 114  
 Db 60663 TCTTCACACAACTTAGATTAGCATTTATGAGTTCTTACTAGCAATTTGCTC 60611

RESULT 22  
 AP003409  
 LOCUS 161818 bp DNA HTG 15-MAR-2001  
 DEFINITION Oryza sativa chromosome 1 clone B1131G08, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in ordered pieces.  
 ACCESSION AP003409  
 VERSION AP003409.1 GI:13366112  
 KEYWORDS HTG; HTGS\_PHASE2.  
 SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:B1131G08.  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.  
 REFERENCE 1 (sites)  
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.  
 TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC  
 clone:B1131G08  
 JOURNAL Published Only in Database (2001) In press  
 REFERENCE 2 (bases 1 to 161818)  
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-MAR-2001) Takuji Sasaki, National Institute of  
 Agricultural Resources, Rice Genome Research Program, Kannondai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail:tsasaki@agr.affrc.go.jp, URL:<http://vgp.dna.affrc.go.jp/>,  
 Tel:81-298-38-7441, Fax:81-298-38-7468)











Db 53767 TTATCTATAAATCTAGCAGCTTAATTGATCTTAATTAGCAGCTTACT 53814

RESULT 25  
AC016271/c DNA HTG 13-JUL-2000

LOCUS AC016271 80065 bp

DEFINITION Homo sapiens clone RP11-19F12, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC016271

VERSION AC016271.2 GI:9123891

KEYWORDS HTG: HTGS\_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 80065)

REFERENCE Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens, clone RP11-19F12

JOURNAL Unpublished

2 (bases 1 to 80065)

REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferraira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McKernan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrison, D., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teste, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced g1:6467063.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3728

Center clone name: 19\_F\_12

-----

\* NOTE: This record contains 82 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

1 903: contig of 903 bp in length

\* 904 1003: gap of 100 bp

\* 1004 1895: contig of 892 bp in length

\* 1896 1995: gap of 100 bp

\* 1996 2890: contig of 895 bp in length

\* 2891 2990: gap of 100 bp

\* 2991 3816: contig of 826 bp in length

\* 3817 3916: gap of 100 bp

\* 3917 4776: contig of 860 bp in length

\* 4777 4876: gap of 100 bp

\* 4877 5746: contig of 870 bp in length

\* 5747 5846: gap of 100 bp

\* 5847 6741: contig of 895 bp in length

\* 6742 6841: gap of 100 bp

\* 6842 7705: contig of 864 bp in length

\* 7706 7805: gap of 100 bp

\* 7806 8688: contig of 883 bp in length

\* 8689 8788: gap of 100 bp

\* 8789 9660: contig of 872 bp in length

\* 9661 9760: gap of 100 bp

\* 9761 10661: contig of 901 bp in length

\* 10662 10761: gap of 100 bp

\* 10762 11625: contig of 864 bp in length

\* 11626 11725: gap of 100 bp

\* 11726 12619: contig of 894 bp in length

\* 12620 12719: gap of 100 bp

\* 12720 13621: contig of 902 bp in length

\* 13622 13721: gap of 100 bp

\* 13722 14562: contig of 861 bp in length

\* 14583 14682: gap of 100 bp

\* 14683 15537: contig of 855 bp in length

\* 15538 15637: gap of 100 bp

\* 15638 16517: contig of 880 bp in length

\* 16518 16617: gap of 100 bp

\* 16618 17504: contig of 887 bp in length

\* 17505 17604: gap of 100 bp

\* 17605 18479: contig of 875 bp in length

\* 18480 18579: gap of 100 bp

\* 18580 19458: contig of 879 bp in length

\* 19459 19558: gap of 100 bp

\* 19559 20428: contig of 870 bp in length

\* 20429 20528: gap of 100 bp

\* 20529 21424: contig of 896 bp in length

\* 21425 21524: gap of 100 bp

\* 21525 22423: contig of 899 bp in length

\* 22424 22523: gap of 100 bp

\* 22524 23367: contig of 844 bp in length

\* 23368 23467: gap of 100 bp

\* 23468 24369: contig of 902 bp in length

\* 24370 24469: gap of 100 bp

\* 24470 25324: contig of 855 bp in length

\* 25325 25424: gap of 100 bp

\* 25425 26267: contig of 843 bp in length

\* 26268 26367: gap of 100 bp

\* 26368 27235: contig of 868 bp in length

\* 27236 27335: gap of 100 bp

\* 27336 28210: contig of 875 bp in length

\* 28211 28310: gap of 100 bp

\* 28311 29200: contig of 890 bp in length

\* 29201 29300: gap of 100 bp

\* 29301 30174: contig of 874 bp in length

\* 30175 30274: gap of 100 bp

\* 30275 31162: contig of 888 bp in length

\* 31163 31262: gap of 100 bp

\* 31263 32105: contig of 843 bp in length

\* 32106 32205: gap of 100 bp

\* 32206 33111: contig of 906 bp in length

\* 33112 33211: gap of 100 bp

\* 33212 34119: contig of 908 bp in length

\* 34120 34219: gap of 100 bp

\* 34220 35103: contig of 884 bp in length

\* 35104 35203: gap of 100 bp

\* 35204 36064: contig of 861 bp in length

\* 36065 36164: gap of 100 bp

\* 36165 37033: contig of 869 bp in length

\* 37034 37133: gap of 100 bp

\* 37134 38021: contig of 888 bp in length

\* 38022 38121: gap of 100 bp

\* 38122 39018: contig of 897 bp in length

\* 39019 39118: gap of 100 bp

\* 39119 40034: contig of 916 bp in length

\* 40035 40134: gap of 100 bp

\* 40135 41047: contig of 913 bp in length

\* 41048 41147: gap of 100 bp

\* 41148 42068: contig of 921 bp in length

\* 42069 42168: gap of 100 bp



```

* 42169 43048: contig of 880 bp in length
* 43049 43148: gap of 100 bp
* 43149 44027: contig of 879 bp in length
* 44028 44127: gap of 100 bp
* 44128 44989: contig of 862 bp in length
* 44990 45089: gap of 100 bp
* 45090 45956: contig of 867 bp in length
* 45957 46056: gap of 100 bp
* 46057 46940: contig of 884 bp in length
* 46941 47040: gap of 100 bp
* 47041 47939: contig of 899 bp in length
* 47940 48039: gap of 100 bp
* 48040 48830: contig of 791 bp in length
* 48831 48930: gap of 100 bp
* 48931 49793: contig of 863 bp in length
* 49794 49893: gap of 100 bp
* 49894 50799: contig of 906 bp in length
* 50800 50899: gap of 100 bp
* 50900 51768: contig of 869 bp in length
* 51769 51868: gap of 100 bp
* 51869 52723: contig of 855 bp in length
* 52724 52823: gap of 100 bp
* 52824 53686: contig of 863 bp in length
* 53687 53786: gap of 100 bp
* 53787 54671: contig of 885 bp in length
* 54672 54771: gap of 100 bp
* 54772 55656: contig of 885 bp in length
* 55657 55756: gap of 100 bp
* 55757 56644: contig of 888 bp in length
* 56645 56744: gap of 100 bp
* 56745 57637: contig of 893 bp in length
* 57638 57737: gap of 100 bp
* 57738 58616: contig of 879 bp in length
* 58617 58716: gap of 100 bp
* 58717 59555: contig of 839 bp in length
* 59556 59655: gap of 100 bp
* 59656 60555: contig of 900 bp in length
* 60556 60655: gap of 100 bp
* 60656 61541: contig of 886 bp in length
* 61542 61641: gap of 100 bp
* 61642 62547: contig of 906 bp in length
* 62548 62647: gap of 100 bp
* 62648 63491: contig of 844 bp in length
* 63492 63591: gap of 100 bp
* 63592 64444: contig of 853 bp in length
* 64445 64544: gap of 100 bp
* 64545 65420: contig of 876 bp in length
* 65421 65520: gap of 100 bp
* 65521 66396: contig of 876 bp in length
* 66397 66496: gap of 100 bp
* 66497 67376: contig of 880 bp in length
* 67377 67476: gap of 100 bp
* 67477 68377: contig of 901 bp in length
* 68378 68477: gap of 100 bp
* 68478 69360: contig of 883 bp in length
* 69361 69460: gap of 100 bp
* 69461 70336: contig of 876 bp in length
* 70337 70436: gap of 100 bp
* 70437 71288: contig of 852 bp in length

```

Query Match 24.1%; Score 29.6; DB 2; Length 80065;  
 Best Local Similarity 53.6%; Pred. No. 16;  
 Matches 59; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```

Oy 3 ggtattgtaaatgtagatgccttacaacccctgatacctaataatgagcgaatgcgaggct 62
Db 12614 GGGCTTCAAAAGTGAAGTATGAACTGGGGAAGCAATGGCACACAAAGATGN 12555
Oy 63 gcaactgactgtaagttcacatgtaagtgatactactcaattat 112
Db 12554 GNACTCTCTGGCGACACTGCACAGCAAGGAGAACATGATCTCAGATGTAT 12505

```

```

RESULT 26
AC005333 94797 bp DNA INV 29-JUL-1998
LOCUS AC005333
DEFINITION Drosophila melanogaster DNA sequence (Pls DS06477 (D116) and
DS03026 (D126)), complete sequence.
ACCESSION AC005333 AC004243 AC001697 AC001698 AC001699 AC001700 AC002577
AC003702 AC001701 AC001702 AC001703 AC001704 AC001705 AC003129
AC001706 AC001707 AC001708 AC001709 AC001710 AC001711 AC002578
AC001712 AC003703 AC003704 AC004271 AC003138 AC003716
AC005333.1 GI:3347820
VERSION HTG.
KEYWORDS Drosophila melanogaster (Subclones in sac from P1 clones DS06477
SOURCE (D116) and DS03026 (D126)) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 94797)
AUTHORS Celniker, S.E., George, R.A., Galle, R.F., Hoskins, R.A.,
Svitskas, R.R., Harris, N.L., Abdayani, A., Arcata, T.T., Baxter, E.,
Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E.,
Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Punch, E., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R.,
Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.
Sequencing of Drosophila chromosome 2L, region 25F3-26A2
Unpublished (1997)
2 (bases 1 to 94797)
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1998) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
AUTHORS Sequence submitted by:
COMMENT Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
email to drosophila@mc.lbl.gov.
Library locations: 45-68, 50-32.
FEATURES
source
1..94797
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/map="25F3-26A2"
/clone="Pls DS06477 (D116) and DS03026 (D126)"
/note="DS06477 (d116) extends from bp 1 to bp 84,543 and
DS03026 (d126) is a bridge extending from bp 83,209 to bp
94,797."
BASE COUNT 27561 a 19822 c 19879 g 27355 t
ORIGIN

```

Query Match 24.1%; Score 29.6; DB 3; Length 94797;  
 Best Local Similarity 54.6%; Pred. No. 16;  
 Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```

Oy 6 attgtaaatgtagatgccttacaacccctgatacctaataatgagcgaatgcgaggctcga 65
Db 11285 ATTGTAATATTAATGATCGTTTCTTAACCTAATGGCAAGGTTTATAGAACTAAACTGAT 11344
Oy 66 ctgactgtaagttcacatgtaagtgatactactcaattat 113

```



Db 11345 TTATCTCATTAAGCTTAAATGATCTTAATGACTTACT 11392

RESULT 27  
AC092239/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC092239 166441 bp DNA INV 30-JUN-2001  
Drosophila melanogaster, chromosome 2L, region 25F-26X, BAC clone  
BACR48E1, complete sequence.  
AC092239  
AC092239.1 GI:14578120  
HTG.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS

Holt, R.A., Evans, C.A., Gocayne, J.D., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busan, D.A., Carlson, J.W., Center, A., Champagne, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferritera, S., Frisze, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Idegawa, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Pihounenavong, S., Piltman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.  
Sequencing of Drosophila chromosome 2L, region 25F-26X  
Unpublished  
2 (bases 1 to 166441)

JOURNAL  
REFERENCE  
AUTHORS

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busan, D.A., Carlson, J.W., Center, A., Champagne, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferritera, S., Frisze, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Idegawa, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Pihounenavong, S., Piltman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.  
Direct Submission  
Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US  
Sequence submitted by:  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

TITLE  
JOURNAL

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgpre@fruitfly.berkeley.edu](mailto:bdgpre@fruitfly.berkeley.edu).

FEATURES  
SOURCE

1. 166441  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="25F-26X"  
/clone="BACR48E14 (D1336)"  
/clone\_11b="RP1-98 (Roswell) Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
PRACe3.6")

BASE COUNT  
47904 a 35803 c 35579 g 47155 t

## ORIGIN

Query Match 24.18; Score 29.6; DB 3; Length 166441;  
Best Local Similarity 54.68; Pred. No. 17;  
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 6 atttaaatgtagatgcctaccacccctgtagccctaatgagcgcaatgagggcgca 65  
Db 85264 ATTGTAATATATTAGATGCTTACTTAAGTACACTTAATGACTTACT 85205

Qy 66 ctgactgtagaattcacatgtaagtgatatacttcaattatc 113  
Db 85204 TTATCTCATTAAGCTTAAATGATCTTAATGACTTACT 85157

RESULT 28  
AE003611/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AE003611 259973 bp DNA INV 04-OCT-2000  
Drosophila melanogaster genomic scaffold 142000013386055 section 4  
of 63, complete sequence.  
AE003611 AE002690  
AE003611.2 GI:10728594  
HTG.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champagne, M., Pfeiffer, B., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos, G.L., Arrill, J.F., Agbayani, A., An, H.J., Andrews, Pfanhko, C., Baldwin, D., Ballew, R.M., Basu, A., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brothier, P., Burris, K.C., Busan, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferriz, C., Ferritera, S., Fleischmann, W., Foster, C., Gabriellian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Idegawa, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Kelchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Laoko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Mlyniska, N.V., Moberly, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Paclet, J.W., Palazolo, M., Piltman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spter, E., Spalding, A.C., Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yen, R.F., Zaveri, J.S., Zhao, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhou, X., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

THE genome sequence of Drosophila melanogaster  
Science 287 (5461), 2185-2193 (2000)  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.  
2 (bases 1 to 259973)



TITLE Direct Submission  
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT On Oct 9, 2000 this sequence version replaced gi:7297008.  
FEATURES  
source 1. 259973  
location/Qualifiers  
organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
join(<16590..16757,16815..17687,17749..>17886)  
/gene="CG7236"  
/product="CT22313"  
/db\_xref="FLYBASE:FBan0007236"  
/db\_xref="FLYBASE:FBgn0031730"  
<16590..>17886  
/gene="CG7236"  
/map="25F4-26A1"  
/db\_xref="FLYBASE:FBan0007236"  
/db\_xref="FLYBASE:FBgn0031730"  
join(16590..16757,16815..17687,17749..17886)  
/gene="CG7236"  
/note="CG7236 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007236"  
/db\_xref="FLYBASE:FBgn0031730"  
/protein\_id="AAF52279.1"  
/db\_xref="GI:7297009"  
/translation="MDREKLSRLGSGYVYKCRDRTGALVAVKRESEDPAL  
RKIALREIRLKNLKNLIVSLVEFRRRRLHVEFCELTVLHELHRRPQCEPL  
TKOICYOTLLGAVCHGKGLCHRDIKPENILITLAGOVYKLDGFEARMISPEENTDY  
VATRWYRDELIVGDTGYCTPDYMAIGLFAELVGRGALMGRSDVDLYLIRKTLG  
DLRPHIOIFGONEFFKGIITLPPPTLEPLEDRKPKSQONPLTIDFKCCLKDPTK  
RWSCKLKHSYFDYIAKQRELEHNSLEANLQOOLASQOFMLATAAOQLGSPA  
QAAIAAARDKSKTSNTSLPLSTQHHNHPQDVKLQPLKNNANLHRTHHNPTI  
"  
complement(join(<19824..20287,20339..>20411))  
/gene="CG14007"  
/product="CT33563"  
/db\_xref="FLYBASE:FBan0014007"  
/db\_xref="FLYBASE:FBgn0031731"  
/evidence=not\_experimental  
complement(<19824..>20411)  
/gene="CG14007"  
/map="25F5-26A1"  
/db\_xref="FLYBASE:FBan0014007"  
/db\_xref="FLYBASE:FBgn0031731"  
/evidence=not\_experimental  
complement(join(19824..20287,20339..20411))  
/gene="CG14007"  
/note="CG14007 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0014007"  
/db\_xref="FLYBASE:FBgn0031731"  
/evidence=not\_experimental  
/protein\_id="AAF52280.1"  
/db\_xref="GI:7297010"  
/translation="MIMLKQIINGLLIITLLGCFYLFVGVNACTGAMFOIGLY  
KSARYCELYGDDPPEHPLAMRSNKDSEKPLVLELMHMDHNDIEEKRSLGILMPR  
ROPVSVLPQAAVPOPLKWMQLQALSTGDDHQPDPSVYNNMHTNELLNCGDVO  
RKSELNRYAENOLHRWY"  
complement(join(<22642..23763,23844..24295))  
/gene="CG11149"  
/product="CT31163"  
/db\_xref="FLYBASE:FBan0011149"  
/db\_xref="FLYBASE:FBgn0031732"  
complement(<22642..>24295)  
/gene="CG11149"  
/map="25F5-26A1"  
/db\_xref="FLYBASE:FBan0011149"  
/db\_xref="FLYBASE:FBgn0031732"  
complement(join(22734..23763,23844..24295))  
/gene="CG11149"  
/note="CG11149 gene product"

/codon\_start=1  
/db\_xref="FLYBASE:FBan0011149"  
/db\_xref="FLYBASE:FBgn0031732"  
/protein\_id="AAF52281.1"  
/db\_xref="GI:7297011"  
/translation="MATSFCNRKRYRLMPVILFENVALLWBNLQAOAKLSSNT  
GGCKSVGLLISNESNPSRFSANFHLANQHKIITRRESATTKELRSLKCRDSSLF  
ERLQGEFEMLQNLVWGRKSRFVGCAESVYTTTNDFTFEDNLKLVVSRKAPVFAI  
HPGYDLNTLTDAIRYVRNCLPESDSDIWVSFHYFENQMPPEYVPEDEAVLMPH  
MCTLANGSLVSPLYQIPIPTDYSKRAMLYPIVNGRIARLATNTHIFACDIELYP  
SGVFQEDLDWYARNHVSVALDPRORRNVPLPEFEITKCAKVPDEDELLATYRQO  
AVFPHLKCPTCHTIIPGQEWLNRTSRADHLLHYFSKALRKRFRAPEFVSDNTEP  
LDEBVTWEGOSNKRIOVSYCSQVSGOCLLONAMCLDIEHYVLSAPFLVHSP  
GIKQSSKADSTRLQYAKETKFIKKIIEPEYRLFGKNSAKT"  
join(25407..25577,25629..27907,27971..28124)  
/gene="slp1"  
/product="CT22315"  
/db\_xref="FLYBASE:FBan0007238"  
/db\_xref="FLYBASE:FBgn0024191"  
<25407..>28124  
/gene="slp1"  
/note="CG7238"  
/map="25F5-26A1"  
/db\_xref="FLYBASE:FBan0007238"  
/db\_xref="FLYBASE:FBgn0024191"  
join(25469..25577,25629..27907,27971..28102)  
/gene="slp1"  
/note="slp1 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007238"  
/db\_xref="FLYBASE:FBgn0024191"  
/protein\_id="AAF52282.1"  
/db\_xref="GI:7297012"  
/translation="MSDNDYERFETDYDLDFNEFININRPGRSRHOQIYIMADSE  
EESGEGGCKRRGRARPKDYTMVNVVAGIGQAGKKKKALQADDEGSGQEKAE  
ADGGEESDASAAGRARGDNDPSSNSSEEPYTLRSKQSTPPQDRSHIASERV  
CAMEPHTGICAKLILQMGYEPGKGLGNDLOGISHPYQVAYHRKRGALGAYGPTAS  
IGGKTNSKIKVDEYREAKFERDQLNKRRKSAGAEEMERGRIYKSVEEYIAGS  
HSGKLSEKLSKRLGNVYIDMTGPEKRSYGHALGQAKITPEETLYDEYERK  
APACVFAMPBELTHNLQILVSOEEOIIAIDNOECSQQALESERKLEIYQLEER  
NHITRLEESLEVERLIDNPDLSPQARLRELLVYAAEFHFGADLAAGVIAL  
LKRELWQOPLNTEPTEPLIKKRWGMQOGDAEQPRNVFDVSSLIWAGVWSPR  
SSAANWQREHPNASSLIDNAPLRLSPVSDIYSEQLYRLVAGVQEMPDTYVPI  
DSWVLPWHAHLIGSKLEAVYPOIRSKGLIALRAVSPDRSRAVLTWQAFPEEQ  
BFLQRYIVPKLOATLGLLIDPMQDLELMQVWEMHELIDPMYMAOLLRHFFPRM  
QVLYVWLNOSPDYAEISRWTGMSLSEPLLRPSYKELHRLALEIMHRASTLLOP  
TYTPPPPPVPAPVIMMDLIHPPAOLEFKEVLSOCCADIGIIPAPLPGRRMGKQIY  
FVGKLCFYCIDRHVCMVSDGSRSPVSLNHLERSGTGL"  
complement(<31796..>33085)  
/gene="CG14006"  
/product="CT33562"  
/db\_xref="FLYBASE:FBan0014006"  
/db\_xref="FLYBASE:FBgn0031733"  
/evidence=not\_experimental  
complement(<31796..>33085)  
/gene="CG14006"  
/map="25F5-26A1"  
/db\_xref="FLYBASE:FBan0014006"  
/db\_xref="FLYBASE:FBgn0031733"  
/evidence=not\_experimental  
complement(31796..33085)  
/gene="CG14006"  
/note="CG14006 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0014006"  
/db\_xref="FLYBASE:FBgn0031733"  
/db\_xref="FLYBASE:FBgn0031733"  
/protein\_id="AAF52283.1"  
/db\_xref="GI:7297013"  
/translation="MMAPLAFVILGLNVAASSFVMDNMFVEQOATYLRPGEPHIO  
GOTLIMLYVNNVOYVCOPTAFSPMNVDSNRLRLQIAAGEDYTOYKQATYRVNH  
AHSQLDCECHKPLGHSADPRIVPAIYEHSCYGIYTSKRPVLTLEVTSFLERS



```

Query Match ..... 24.1% Score 29.6; DB 3; Length 259973;
Best Local Similarity 54.6%; Pred. No. 17;
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Oy      6 attgtaaatgtgagatgccttaccaaccctgatgacctaatggcgaatgcaggcgctgca 65
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210464 ATTGTAATATTATAGATCGCTTCTCTAACTAATGGCAACAGTTTAAATAGAACATAACTGTAT 210405

Oy      66 ctgactggtaagtattccacatgtaagtgatactattcttcaatttcatt 113
        || || || || || || || || || || || || || || || || || || || || ||
Db 210404 TTATCTCATTAACAATTAGCATCTAAATATGATTAATTAATAGCACTTACT 210357

RESULT 29
AC021110
LOCUS Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
DEFINITION AC021110
ACCESSION AC021110.3 GI:8439978
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 155172)
TITLE Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 155172)
TITLE Waterston,R.H.
JOURNAL Direct Submission
SUBMITTER Submitted (14-JAN-2000) Genome Sequencing Center, Washington
UNIVERSITY University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 10, 2000 this sequence version replaced gi:7021826.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0238P13
----- Summary Statistics -----
Sequencing vector: M13; 73%
Sequencing vector: plasmid; 27%
Chemistry: Dye-primer ET; 73% of reads
Chemistry: Dye-terminator Big Dye; 27% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151162 bases at least Q40
Consensus quality: 153249 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 154672; sum-of-contigs
Quality coverage: 4.08 in Q20 bases; agarose-fp
Quality coverage: 4.69 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 6 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1          9265: contig of 9265 bp in length
*
*          9266          9365: gap of unknown length
*
*          9366          26504: contig of 17139 bp in length
*
*          26505          26604: gap of unknown length
*
*          44301          44300: contig of 17696 bp in length
*
*          44401          44400: gap of unknown length
*
*          62250          62250: contig of 17850 bp in length
*
*          62251          62250: gap of unknown length

```

FEATURES	source	Location/Qualifiers
*	62351	105744: contig of 4394 bp in length
*	105745	105844: gap of unknown length
*	105845	155172: contig of 49328 bp in length.
		Location/Qualifiers
		1..155172
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="18"
		/clone="RP11-238P13"
		1..9265
	misc_feature	/note="assembly_name:Contig1"
		9366..26504
	misc_feature	/note="assembly_name:Contig2"
		26605..44300
	misc_feature	/note="assembly_name:Contig3
		clone_end:17
		vector_side:right"
	misc_feature	44401..62250
		/note="assembly_name:Contig4"
	misc_feature	62351..105744
		/note="assembly_name:Contig5"
	misc_feature	105845..155172
		/note="assembly_name:Contig6"
	BASE COUNT	51543 a 28385 c 28043 g 46698 t 503 others
	ORIGIN	
Query Match	23.9%	Score 29.4; DB 2; Length 155172;
Best Local Similarity	55.3%	Pred. No. 19;
Matches	57; Conservative	0; Mismatches 46; Indels 0; Gaps 0;
Qy	18	agatgccttaacaacctatgacctaatgagcgcaatgcgagggctgcactgactgtaag 77
Db	83677	AGGTCCTTTANGCACTTCATTAGCTTATVAGAAAATTTGAATGCTGCACATATTTGTAAAT 83736
Qy	78	tttccatgtaagtgatgatacttcttcaattatctccacaca 120
Db	83737	TACAGTATGTAACTGTGCTGGAACCTTTCTTACTTAAACGA 83779
RESULT	30	
AL359922		
LOCUS	AL359922	168656 bp DNA PRI 17-JAN-2001
DEFINITION	Human DNA sequence from clone RP11-70L8 on chromosome 9, complete sequence.	
ACCESSION	AL359922	
VERSION	AL359922.10	GI:12191425
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
REFERENCE	1 (bases 1 to 168656)	
AUTHORS	Laird,G.	
TITLE	Direct Submission	
JOURNAL	Submitted (17-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquereyes@sanger.ac.uk Clone	
COMMENT	requests: clonerequests@sanger.ac.uk	
	On Jan 13, 2001 this sequence version replaced gi:11878000.	
	During sequence assembly data is compared from overlapping clones.	
	Where differences are found these are annotated as variations	
	together with a note of the overlapping clone name. Note that the	
	variation annotation may not be found in the sequence submission	
	corresponding to the overlapping clone, as we submit sequences with	
	only a small overlap as described above.	
	This sequence has been finished according to sequence map criteria	
	as follows. An attempt is made to resolve all sequencing problems,	
	such as compressions and repeats, but not necessarily within known	
	annotated repeat sequence elements. Where the sequence is	
	ambiguous, there is an annotation using the "unsure" feature key.	
	The following abbreviations are used to associate primary accession	
	numbers given in the feature table with their source databases:	
	Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information	



on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 Chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Ch9>  
 RP11-70L8 is from the library RP11-11.1 constructed by the group of  
 Plier de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-70L8 The true left  
 end of clone RP11-14912 is at 118787 in this sequence. The true  
 right end of clone RP11-47303 is at 73731 in this sequence.

## FEATURES

Source

1. 168656

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="9"

/clone\_11b="RP11-70L8"

/clone\_11b="RP11-11.1"

/note="Single clone region. Assembly confirmed by  
 restriction digest data."

/note="Single clone region. Assembly confirmed by  
 restriction digest data."

/note="Single clone region. Assembly confirmed by  
 restriction digest data."

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

/note="Single clone region. Assembly confirmed by  
 restriction digest data"

## COMMENT

On May 30, 2000 this sequence version replaced gi:7259325.

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: [hattori@gsc.riken.go.jp](mailto:hattori@gsc.riken.go.jp)

----- Project Information

Center project name: HumDrafl18

Center clone name: RP11-635B11

----- Summary Statistics

Sequencing vector: PCR products: 100% of reads

Chemistry: Dye-terminator PCR products: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 159194 bases at least Q40

Consensus quality: 167087 bases at least Q30

Consensus quality: 170693 bases at least Q20

Insert size: 172815; sum-of-contigs

Quality coverage: 4.94x in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently consists of  
 21 contigs. The true order of the pieces is not known and their  
 order in this sequence record is arbitrary. Gaps between the  
 contigs are represented as runs N, but the exact sizes of the gaps  
 are unknown. This record will be updated with the finished sequence  
 as soon as it is available and the accession number will be  
 preserved

1 58823 contig of 58823 bp in length

58924 73243 contig of 14320 bp in length

73344 83572 contig of 10229 bp in length

83673 94547 contig of 10875 bp in length

94648 104597 contig of 9950 bp in length

104698 114347 contig of 9650 bp in length

114448 122080 contig of 7633 bp in length

122181 128465 contig of 6265 bp in length

128566 135791 contig of 7226 bp in length

135892 142934 contig of 7043 bp in length

143035 147999 contig of 4965 bp in length

148100 151734 contig of 3633 bp in length

151835 155576 contig of 3742 bp in length

155677 159193 contig of 3517 bp in length

159294 161359 contig of 2056 bp in length

161460 163515 contig of 2056 bp in length

163616 166614 contig of 2999 bp in length

166715 169056 contig of 2342 bp in length

169157 171527 contig of 2371 bp in length

171628 173248 contig of 1621 bp in length

173349 174815 contig of 1467 bp in length

Sequence updated (26-May-2000).

NOTE: This is a 'working draft' sequence. It currently  
 consists of 21 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 58823: contig of 58823 bp in length

58824 58923: gap of 100 bp

58924 73243: contig of 14320 bp in length

73244 73343: gap of 100 bp

73344 83572: contig of 10229 bp in length

83573 83672: gap of 100 bp

83673 94547: contig of 10875 bp in length

94548 94647: gap of 100 bp

94648 104597: contig of 9950 bp in length

104598 104697: gap of 100 bp

104698 114347: contig of 9650 bp in length

114348 114447: gap of 100 bp

114448 122080: contig of 7633 bp in length

122081 122180: gap of 100 bp

122181 128465: contig of 6285 bp in length

128466 128565: gap of 100 bp

128566 135791: contig of 7226 bp in length

RESULT 31  
 AP001390 174815 bp DNA HTG 30-MAY-2000  
 LOCUS  
 DEFINITION Homo sapiens chromosome 18 clone RP11-635B11 map 18q23, WORKING  
 DRAFT SEQUENCE, 21 unordered pieces.  
 ACCESSION AP001390  
 VERSION AP001390.2 GI:8117301  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRIFT.  
 SOURCE Homo sapiens DNA, clone:RP11-635B11.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 174815)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,  
 Homo sapiens 174,815 genomic DNA of 18q23  
 Published Only in Database (2000) In press  
 2 (bases 1 to 174815)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,  
 Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
 Japan (E-mail:hattori@gsc.riken.go.jp,  
 URL:<http://hgp.gsc.riken.go.jp/>, Tel:81-42-778-9923,  
 Fax:81-42-778-9924)



FEATURES	source
*	135792 135891: gap of 100 bp
*	135892 142834: contig of 7043 bp in length
*	142835 143034: gap of 100 bp
*	143035 147999: contig of 4665 bp in length
*	148000 148099: gap of 100 bp
*	148100 151734: contig of 3635 bp in length
*	151735 151834: gap of 100 bp
*	151835 155576: contig of 3742 bp in length
*	155577 155676: gap of 100 bp
*	155677 159193: contig of 3517 bp in length
*	159194 159293: gap of 100 bp
*	159294 161359: contig of 2066 bp in length
*	161360 161459: gap of 100 bp
*	161460 163515: contig of 2056 bp in length
*	163516 163615: gap of 100 bp
*	163616 166614: contig of 2899 bp in length
*	166615 166714: gap of 100 bp
*	166715 169056: contig of 2342 bp in length
*	169057 169156: gap of 100 bp
*	169157 171527: contig of 2371 bp in length
*	171528 171627: gap of 100 bp
*	171628 173248: contig of 1621 bp in length
*	173249 173348: gap of 100 bp
*	173349 174815: contig of 1467 bp in length
	Location/Qualifiers
	1..174815

```

/misc_feature      /organism="Homo sapiens"
/misc_feature      /db_xref="taxon:9606"
/misc_feature      /chromosome="18"
/misc_feature      /map="18q23"
/misc_feature      /clone="RP11-635B11"
/misc_feature      1..58823
/misc_feature      /note="assembly_fragment"
/misc_feature      58924..73343
/misc_feature      /note="assembly_fragment"
/misc_feature      73344..83572
/misc_feature      /note="assembly_fragment"
/misc_feature      83673..94547
/misc_feature      /note="assembly_fragment"
/misc_feature      94648..104597
/misc_feature      /note="assembly_fragment"
/misc_feature      104698..114347
/misc_feature      /note="assembly_fragment"
/misc_feature      114448..122050
/misc_feature      /note="assembly_fragment"
/misc_feature      122181..128465
/misc_feature      /note="assembly_fragment"
/misc_feature      128566..135791
/misc_feature      /note="assembly_fragment"
/misc_feature      135892..142934
/misc_feature      /note="assembly_fragment"
/misc_feature      143035..147999
/misc_feature      /note="assembly_fragment"
/misc_feature      148100..151734
/misc_feature      /note="assembly_fragment"
/misc_feature      151835..155576
/misc_feature      /note="assembly_fragment"
/misc_feature      153677..159193
/misc_feature      /note="assembly_fragment"
/misc_feature      159294..161359
/misc_feature      /note="assembly_fragment"
/misc_feature      161460..163515
/misc_feature      /note="assembly_fragment"
/misc_feature      163616..166614
/misc_feature      /note="assembly_fragment"
/misc_feature      166715..169056
/misc_feature      /note="assembly_fragment"
/misc_feature      169157..171527
/misc_feature      /note="assembly_fragment"
/misc_feature      171628..173248
/misc_feature      /note="assembly_fragment"
/misc_feature      173349..174815
/misc_feature      /note="assembly_fragment"

```

BASE COUNT	52310 a	30923 c	32183 g	57399 t	2000 others
ORIGIN					
Query Match	23.9%;	Score 29.4;	DB 2;	Length 174815;	
Best Local Similarity	55.3%;	Pred. No. 19;			
Matches	57;	Conservative	0;	Mismatches	46; Indels 0; Gaps 0;
OY	18 agatgaccttacacaccctgatgtacctaatgycgcgaatcgagggctgcacttgtagtaag	77			
Db	71924 AGGTCCTTTATGCACCTTCATTAGCCTTATAGAAAATAATTAATTCGTGCACAATATATGTAAAT	71983			
OY	78 ttcaacatgaaggatgatactaattcatttcacaaca	120			
Db	71984 TACAGTAGTAACTGTGGCTGAACCCTTTCTTACTAACCA	72026			

RESULT	32	
AC091589		
LOCUS	AC091589	178056 bp DNA
DEFINITION	Homo sapiens chromosome 18 clone RP11-635B11 map 18, *** SEQUENCING	26-JUL-2001
	IN PROGRESS ***, 2 ordered pieces.	
ACCESSION	AC091589	
VERSION	AC091589.6	GI:15022066
KEYWORDS	HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE	1 (bases 1 to 178056)
AUTHORS	Bitren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens chromosome 18, clone RP11-635B11
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 178056)
AUTHORS	Bitren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,

AUTHORS

Barrett, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barua, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Cheepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Deatrelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hesford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karlas, A., LaRoque, K., Lamaszates, R., Landers, T., Lehoczeky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nordu, C., Norman, C., H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, U., Peterson, K., Phunhahang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnuppach, R., Seaman, S., Severy, P., Sounguez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, J., Zody, M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 26, 2001 this sequence version replaced gi:14994291. All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center

Center Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information  
 Center project name: L13182  
 Center clone name: 635\_B\_11

---

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. Gaps between the contigs







Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>  
VECTOR: pcypac2  
IMPORTANT: This sequence is not the entire insert of clone RPl-95121. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RPl-95121 is at 120515 in this sequence. The true left end of clone RPl-95121 is at 4877 in this sequence. The true right end of clone RPl-242N11 is at 100 in this sequence.

## FEATURES

source  
1. .120515  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RPl-95121"  
/clone\_lib="RPCT-1"  
1. .62  
repeat\_region  
/note="31 copies 2 mer at 71% conserved"  
complement(640..1130)  
/note="match: GSS: Em:AQ595695"  
751..989  
repeat\_region  
/note="AluIo repeat: matches 59. .297 of consensus"  
1620..2460  
/note="LIM4 repeat: matches 4616. .5481 of consensus"  
2472..3677  
/note="LIMC repeat: matches 1047. .2247 of consensus"  
3751..3929  
/note="AluSg/x repeat: matches 134. .310 of consensus"  
3932..4089  
/note="LIMC repeat: matches 890. .1062 of consensus"  
4090..4469  
/note="match: GSS: Em:AQ013144"  
4570..4746  
/note="MT1A2 repeat: matches 200. .374 of consensus"  
4824..4931  
/note="MT1A2 repeat: matches 55. .162 of consensus"  
5001..5050  
/note="25 copies 2 mer aa 72% conserved"  
complement(5063..5482)  
/note="match: GSS: Em:AQ511994"  
5279..5376  
/note="L2 repeat: matches 2575. .2678 of consensus"  
5507..5951  
/note="match: GSS: Em:AQ526871"  
5515..6122  
/note="match: GSS: Em:AQ530256"  
complement(5957..6576)  
/note="match: GSS: Em:AQ630441"  
complement(6082..6597)  
/note="match: GSS: Em:AQ693789"  
6095..6259  
/note="MIR repeat: matches 5. .180 of consensus"  
6362..6475  
/note="MER5A repeat: matches 50. .161 of consensus"  
6591..7153  
/note="match: GSS: Em:AQ379073"  
6786..6934  
/note="MER5B repeat: matches 1. .173 of consensus"  
7319..7428  
/note="L2 repeat: matches 2575. .2702 of consensus"  
7609..7640  
/note="16 copies 2 mer ac 87% conserved"  
complement(10277..10730)  
/note="match: GSS: Em:AQ839431"  
10672..11208  
/note="match: GSS: Em:AQ716569"  
10853..11289  
/note="match: GSS: Em:AQ082251"  
12276..12566  
/note="AluIo repeat: matches 1. .298 of consensus"  
12890..13106  
/note="MIR repeat: matches 33. .232 of consensus"

misc\_feature  
complement(12939..13175)  
/note="match: GSS: Em:B85743"  
13214..13514  
/note="match: GSS: Em:AQ429929"  
13216..13542  
/note="match: GSS: Em:AQ835474"  
13516..13551  
/note="18 copies 2 mer tg 100% conserved"  
13518..13695  
/note="match: GSS: Em:AQ429929"  
14150..14378  
/note="MIR repeat: matches 14. .241 of consensus"  
15637..15675  
/note="MADE1 repeat: matches 1. .39 of consensus"  
15901..16030  
/note="MER5A repeat: matches 9. .182 of consensus"  
16173..16202  
/note="15 copies 2 mer gt 96% conserved"  
17819..18027  
/note="MIR repeat: matches 7. .225 of consensus"  
18516..18938  
/note="MT2CB repeat: matches 1. .450 of consensus"  
18934..19450  
/note="MT2D repeat: matches 1. .504 of consensus"  
19465..19521  
/note="MT2D repeat: matches 499. .553 of consensus"  
20072..20270  
/note="LIMB5 repeat: matches 5936. .6171 of consensus"  
20444..24195  
/note="LIM4 repeat: matches 2448. .6279 of consensus"  
25276..25640  
/note="MT1A1 repeat: matches 1. .365 of consensus"  
26400..26653  
/note="AluIo repeat: matches 1. .287 of consensus"  
30430..30896  
/note="L1PA11 repeat: matches 5694. .6165 of consensus"  
33041..33473  
/note="MT2FB repeat: matches 1. .414 of consensus"  
complement(34176..34667)  
/note="match: GSS: Em:AQ111813"  
34344..34474  
/note="LIMC4 repeat: matches 7851. .7977 of consensus"  
34475..34502  
/note="14 copies 2 mer ct 92% conserved"  
34504..34733  
/note="LIMC4 repeat: matches 7625. .7857 of consensus"  
35199..35312  
/note="57 copies 2 mer tt 62% conserved"  
35336..35696  
/note="NAL1 repeat: matches 1170. .1526 of consensus"  
35971..36416  
/note="match: GSS: Em:AQ152633"  
36011..36113  
/note="L1PA15 repeat: matches 1910. .2012 of consensus"  
36115..36503  
/note="MER5B repeat: matches 1. .399 of consensus"  
36513..38620  
/note="MER57-internal repeat: matches 1. .2125 of consensus"  
38624..38694  
/note="AluJ/FRAM repeat: matches 228. .295 of consensus"  
38697..38999  
/note="AluSx repeat: matches 1. .300 of consensus"  
39000..39345  
/note="MER57-internal repeat: matches 2129. .2596 of consensus"  
39339..43921  
/note="MER57-internal repeat: matches 2874. .7531 of consensus"  
43927..44325  
/note="MER57B repeat: matches 1. .403 of consensus"  
44326..46056  
/note="L1PA15 repeat: matches 2010. .3753 of consensus"



```

repeat_region 46057..46329
/note="AluSp repeat: matches 17..290 of consensus"
repeat_region 46310..48228
/note="L1PA15 repeat: matches 3753..6155 of consensus"
repeat_region 48229..48888
/note="L1PA15 repeat: matches 5499..6153 of consensus"
repeat_region 49295..49393
/note="MIR repeat: matches 53..150 of consensus"
repeat_region 49721..49980
/note="L1M4 repeat: matches 2110..2375 of consensus"
misc_feature complement(50308..50438)
/note="match: GSS: Em: A0894428"
misc_feature complement(50343..50444)
/note="match: STS: Em: L17929"
repeat_region 51041..51106

Query Match 23.7%: Score 29.2; DB 9; Length 120515;
Best Local Similarity 57.8%; Pred. No. 22;
Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

OY 18 agatgcctacaacccctgactgaatgacgaatgagggctgactgactgtaag 77
DB 5308 AGAGGCATTCAAAACAAATACCTGCTCTCTCGGAGAAACACAGCTGTGAGG 5367
OY 78 ttcaacatgaagtgtactatcttcaa 107
DB 5368 AGAGGAAAAAGAAATACATATTACAA 5397

RESULT 35
AC013621/c AC013621 159073 bp DNA HTG 20-SEP-2000
LOCUS Homo sapiens clone RP11-11123, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION pieces.
AC013621
VERSION AC013621.5 GI:10198425
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 159073)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-11123
2 (bases 1 to 159073)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Fitzgerald,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karacas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,D., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Traflet,S., Turrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,W.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:9123835.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

```

Center project name: L3173
Center clone name: 11_123
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156407 bases at least Q40
Consensus quality: 158013 bases at least Q30
Consensus quality: 158490 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 158673; sum-of-contigs
Quality coverage: 5.6 in Q20 bases; agarose-fp
Quality coverage: 5.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 30588: contig of 30588 bp in length
* 30589 30688: gap of 100 bp
* 30689 82042: contig of 51354 bp in length
* 82043 82142: gap of 100 bp
* 82143 108087: contig of 25945 bp in length
* 108088 108187: gap of 100 bp
* 108188 134533: contig of 26346 bp in length
* 134534 134633: gap of 100 bp
* 134634 159073: contig of 24440 bp in length.
Location/Qualifiers
1..159073
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-11123"
/clone_lib="RPCT-11 Human Male BAC"
1..30588
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature 30689..82042
/note="assembly-fragment"
82143..108087
/note="assembly-fragment"
108188..134533
/note="assembly-fragment"
134634..159073
/note="assembly-fragment"
clone_end:T7
vector_side:right"
BASE COUNT 52902 a 27848 c 26048 g 51874 t 401 others
ORIGIN
OY 10 taatgtgagatgcctcaacccctgactgaatgacgaatgagggctgactgacga 69
DB 4206 TAAATATCATGATATGACCCCTCTTCATCTTCATTTATTTAGTGTCTGAGCTTATGTC 4147
OY 70 ctggtgaatttcacatgtaagtgaattatctcaattatccacacaca 123
DB 4146 CTGATATCTATCTTGAATAATTATATAGACTTCAACCTAGTCACATGAAA 4093

RESULT 36
AC022863 AC022863 159150 bp DNA HTG 17-MAR-2000
LOCUS Homo sapiens chromosome 4 clone RP11-352124 map 4, WORKING DRAFT
DEFINITION SEQUENCE, 5 unordered pieces.

```



AC022863 GI:7259752  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 159150)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 4, clone RP11-352I24  
 Unpublished  
 2 (bases 1 to 159150)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
 Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G., Castle,A.,  
 Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,  
 Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Leoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
 MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPeeters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
 Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 17, 2000 this sequence version replaced gi:6980325.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L6194  
 Center clone name: 352.I.24  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 155967 bases at least Q40  
 Consensus quality: 157848 bases at least Q30  
 Consensus quality: 158421 bases at least Q20  
 Insert size: 15700; agarose-1p  
 Insert size: 158750; sum-of-contrigs  
 Quality coverage: 6.6 in Q20 bases; agarose-1p  
 Quality coverage: 6.5 in Q20 bases; sum-of-contrigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contrigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 3449: contig of 3449 bp in length  
 \* 3450 3549: gap of 100 bp  
 \* 3550 11150: contig of 7601 bp in length  
 \* 11151 11250: gap of 100 bp  
 \* 11251 48389: contig of 37139 bp in length  
 \* 48390 48489: gap of 100 bp  
 \* 48490 102960: contig of 54471 bp in length  
 \* 102961 105060: gap of 100 bp  
 \* 105061 159150: contig of 56090 bp in length.

FEATURES  
 source  
 Location/Qualifiers  
 1..159150  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-352I24"  
 /clone\_1lb="RP11-11 Human Male BAC"  
 1..3449  
 /note="assembly-fragment"  
 3550..11150  
 /note="assembly-fragment"  
 clone\_end:SP6  
 vector\_side:left"  
 11251..48389  
 /note="assembly-fragment"  
 clone\_end:17  
 vector\_side:left"  
 48490..102960  
 /note="assembly-fragment"  
 103061..159150  
 /note="assembly-fragment"  
 BASE COUNT 50579 a 29976 c 29236 g 48959 t 400 others  
 ORIGIN  
 Query Match 23.7%; Score 29.2; DB 2; Length 159150;  
 Best Local Similarity 56.1%; Pred. No. 23;  
 Matches 55; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
 Qy 26 tacaaccctgactatgcgcacatgcgagatgcgcacatgcgagatgcacat 85  
 Db 13716 TAAACAAATCTTTATTAATCATGCAACAGACGCTTTACTTACTTACTTAC 13775  
 Qy 86 gtaagtgatctactcttcacattatccacacaca 123  
 Db 13776 CTATCGTATATGCACCTTCTATTATTCGCGGTACCA 13813  
 RESULT 37  
 AC021559  
 LOCUS  
 DEFINITION  
 AC021559  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 170022)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 4, clone RP11-753I19  
 Unpublished  
 2 (bases 1 to 170022)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
 Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G., Castle,A.,  
 Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,  
 Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Leoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
 MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPeeters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
 Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.



```

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (16-JUN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 23, 2000 this sequence version replaced 91:6814977.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
-----
Center project name: L5673
Center clone name: 753-L-19
-----
Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161482 bases at least Q40
Consensus quality: 165433 bases at least Q30
Consensus quality: 167439 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 169122; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
* 1 3847: contig of 3847 bp in length
* 3848 3947: gap of 100 bp
* 3948 12621: contig of 8674 bp in length
* 12622 12721: gap of 100 bp
* 12722 23152: contig of 10431 bp in length
* 23153 23252: gap of 100 bp
* 23253 37025: contig of 13773 bp in length
* 37026 37125: gap of 100 bp
* 37126 52252: contig of 15127 bp in length
* 52253 52352: gap of 100 bp
* 52353 66785: contig of 14433 bp in length
* 66786 66885: gap of 100 bp
* 66886 87090: contig of 20205 bp in length
* 87091 87190: gap of 100 bp
* 87191 108749: contig of 21559 bp in length
* 108750 108849: gap of 100 bp
* 108850 142086: contig of 33237 bp in length
* 142087 142186: gap of 100 bp
* 142187 170022: contig of 27836 bp in length.
*
* Location/Qualifiers
* 1. 170022
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="4"
* /map="4"
* /clone="RP11-753L19"
* /clone_lib="RP11 Human Male BAC"
* 1. 3847
* /note="assembly_fragment"
* 3948. 12621
* /note="assembly_fragment"
* 12722. 23152
* /note="assembly_fragment"
* 23253. 37025
* /note="assembly_fragment"
* 37126. 52252
* /note="assembly_fragment"
* 52353. 66785

```

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
misc_feature		/note="assembly_fragment clone_end:SP6 vector_side:right"			
misc_feature		/note="assembly_fragment" 66866..87090			
misc_feature		87191..108749			
misc_feature		/note="assembly_fragment" 108850..142086			
misc_feature		/note="assembly_fragment" 142187..170022			
misc_feature		/note="assembly_fragment clone_end:T7 vector_side:right"			
BASE COUNT	57166 a	29140 c	27735 g	55077 t	904 others
ORIGIN					
Query Match	23.7%	Score 29.2;	DB 2;	Length 170022;	
Best Local Similarity	53.5%	Pred. No.23;			
Matches 61;	Conservative	0;	Mismatches 53;	Indels 0;	Gaps 0;
Qy	10	taaatgtgagatgccttacaccctgatctacctaagtcgacgaagcgtgacactga 69			
Db	30128	TTAAATTCATGCTGATATGACCCCTCTCTTCATCTCTCACTTTATGTCGTACGCTTATGTC 30187			
Qy	70	cttgtaagattcaacatgtaaggtagaacttatcttcaattatgcacacaaca 123			
Db	30188	CTGGTATCTATTACTTGAAATTTATATATAGACTTCACCTGATCCACATGAAA 30241			
RESULT 38					
AC068203/C					
LOCUS	AC068203	181127 bp	DNA	HTG	25-JUN-2000
DEFINITION		Homo sapiens chromosome 4 clone RP11-683C4 map 4,			WORKING DRAFT
VERSION	AC068203				
KEYWORDS	AC068203.2	GI:8705048			
SOURCE		HTG; HTGS_PHASE1; HTGS_DRAFT.			
ORGANISM		human.			
		Homo sapiens			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
		1 (bases 1 to 181127)			
		Birren,B., Linton,L., Nusbaum,C. and Lander,E.			
		2 (bases 1 to 181127)			
		Birren,B., Linton,L., Nusbaum,C. and Lander,E.			
		2 (bases 1 to 181127)			
		Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,			
		Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,			
		Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,			
		Campomali,A., Castle,A., Chopey,I., Colangelo,M., Collins,S.,			
		Collamore,A., Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S.,			
		Dodge,S., Donno,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,			
		Gallagan,J., Gardyna,S., Ginde,S., Goyette,W., Graham,L.,			
		Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,			
		Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,			
		Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczy,J.,			
		Levine,R., Lien,C., Liu,G., Locke,K., MacDonald,P., Margis,N.,			
		McCarthy,M., McEwan,P., McCutk,A., McKernan,K., McPheters,R.,			
		Meldrum,T., Meneus,L., Minova,T., Miranda,C., Mlenka,V., Morrow,J.,			
		Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,			
		O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,			
		Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,			
		Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,			
		Stange-Thomann,N., Stojanovic,N., Sudriamanian,A., Talams,J.,			
		Testaye,S., Theodore,J., Tittell,A., Travers,M., Triggillo,J.,			
		Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,			
		Young,G., Zainoun,J., Zimmer,A. and Zody,M.			
		Direct Submission			
		Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome			
		Research, 320 Charles Street, Cambridge, MA 02141, USA			
		On Jun 25, 2000 this sequence version replaced gi:7671263.			
		All repeats were identified using RepeatMasker:			



Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L10034

Center clone name: 683\_C4

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731.

Consensus quality: 172044 bases at least Q40

Consensus quality: 176527 bases at least Q30

Consensus quality: 178383 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 179427; sum-of-ctdigs

Quality coverage: 4.7 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 bases; sum-of-ctdigs

----- NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 18 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1      2227: contig of 2227 bp in length
*      2228 2327: gap of 100 bp
*      2328 5041: contig of 2714 bp in length
*      5042 5141: gap of 100 bp
*      5142 8953: contig of 3812 bp in length
*      8954 9053: gap of 100 bp
*      9054 13203: contig of 4150 bp in length
*      13204 18520: contig of 5217 bp in length
*      18521 24235: contig of 5615 bp in length
*      24236 24335: gap of 100 bp
*      24336 31130: contig of 6795 bp in length
*      31131 31230: gap of 100 bp
*      31231 38604: contig of 7374 bp in length
*      38605 38704: gap of 100 bp
*      38705 45458: contig of 6754 bp in length
*      45459 45558: gap of 100 bp
*      45559 52091: contig of 6533 bp in length
*      52092 52191: gap of 100 bp
*      52192 58656: contig of 6465 bp in length
*      58657 58756: gap of 100 bp
*      58757 68155: contig of 9399 bp in length
*      68156 68255: gap of 100 bp
*      68256 77655: contig of 9400 bp in length
*      77656 77755: gap of 100 bp
*      77756 85517: contig of 7762 bp in length
*      85518 85617: gap of 100 bp
*      85618 99338: contig of 13721 bp in length
*      99339 99438: gap of 100 bp
*      99439 117256: contig of 17818 bp in length
*      117257 117356: gap of 100 bp
*      117357 138934: contig of 21578 bp in length
*      138935 139034: gap of 100 bp
*      139035 181127: contig of 42093 bp in length.

```

#### FEATURES

source

```

1. 181127
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone_lib="RP11-683C4"
/clone_lib="RP11 Human Male BAC"

```

```

misc_feature 1..2227
             /note="assembly_fragment"
misc_feature 2328..5041
             /note="assembly_fragment"
misc_feature 5142..8953
             /note="assembly_fragment"
misc_feature 9054..13203
             /note="assembly_fragment"
misc_feature 13304..18520
             /note="assembly_fragment"
misc_feature 18621..24235
             /note="assembly_fragment"
misc_feature 24336..31130
             /note="assembly_fragment"
misc_feature 31231..38604
             /note="assembly_fragment"
misc_feature 38705..45458
             /note="assembly_fragment"
misc_feature 45559..52091
             /note="assembly_fragment"
misc_feature 52192..58656
             /note="assembly_fragment"
misc_feature 58757..68155
             /note="assembly_fragment"
misc_feature 68256..77655
             /note="assembly_fragment"
misc_feature 77756..85517
             /note="assembly_fragment"
misc_feature 85618..99338
             /note="assembly_fragment"
misc_feature 99439..117256
             /note="assembly_fragment"
misc_feature 117357..138934
             /note="assembly_fragment"
misc_feature 139035..181127
             /note="assembly_fragment"
             clone_end:17
             vector_side:left"

```

BASE COUNT 53993 a 33116 c 33303 g 59013 t 1702 others

ORIGIN

Query Match 23.7%; Score 29.2; DB: 2; Length 181127;  
 Best Local Similarity 56.1%; Pred. No. 23;  
 Matches 55; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```

QY 26 tacaacctgatgacctaaagcgcaatgcgaagcgcacgtgtaagttcacat 85
   || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 51026 TAAAGCAATCTTTATTAATCATGCAAGAGAGAGCTTTACTTCTTAACCTACAGAA 50967
   || || || | | | | | | | | | | | | | | | | | | | | | | | |
QY 86 gtaagtgatactatcttcaattatccacacaaca 123
   || || || | | | | | | | | | | | | | | | | | | | | | | | |
Db 50966 CTATCGATATGACAGCTCTATTATTCGCGGTAAACA 50929

```

```

RESULT 39
AC023187 181542 bp DNA HTG 01-MAR-2000
LOCUS Homo sapiens chromosome 4 clone RP11-102H15 map 4, WORKING DRAFT
DEFINITION AC023187
SEQUENCE AC023187
ACCESSION AC023187.2 GI:7139717
VERSION 1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181542)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-102H15
JOURNAL Unpublished

```



REFERENCE  
AUTHORS

2 (bases 1 to 181542)  
 Birtten, B., Linton, L., Nusbbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barua, N., Beckerly, R., Beda, F.,  
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,  
 Chesepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,  
 DeCarliano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J.,  
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
 Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
 Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,  
 McSheehers, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, M., Peterson, K.,  
 Pleyer, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Testaf, S., Theodore, J.,  
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
 Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
 Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 1, 2000 this sequence version replaced g1:6957798.  
 All repeats were identified using RepeatMasker:  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 Smit, A.F.A. & Green, P. (1996-1997)

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Center project name: 102\_H15

Center clone name: 102\_H15

Sequencing vector: M13: M77815: 100% of reads

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 173654 bases at least Q40

Consensus quality: 176417 bases at least Q20

Insert size: 182000; agarose-fp

Quality coverage: 3.9 in Q20 bases; agarose-fp

Quality coverage: 3.9 in Q20 bases; sum-of-ctigs

NOTE: This is a 'working draft' sequence. It currently

consists of 23 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1096: contig of 1096 bp in length  
 1097 1196: gap of 100 bp  
 1197 3560: contig of 2364 bp in length  
 3561 3660: gap of 100 bp  
 3661 5983: contig of 2323 bp in length  
 5984 6083: gap of 100 bp  
 6084 8818: contig of 2735 bp in length  
 8819 8918: gap of 100 bp  
 8919 11436: contig of 2518 bp in length  
 11437 11536: gap of 100 bp  
 11537 14541: contig of 3005 bp in length  
 14542 14641: gap of 100 bp  
 14642 18297: contig of 3656 bp in length  
 18298 18397: gap of 100 bp  
 18398 22136: contig of 3739 bp in length  
 22137 22236: gap of 100 bp  
 22237 25665: contig of 3429 bp in length  
 25666 25765: gap of 100 bp  
 25766 30336: contig of 4571 bp in length  
 30337 30436: gap of 100 bp

## FEATURES

## source

30437 34701: contig of 4265 bp in length  
 34702 34801: gap of 100 bp  
 34802 38748: contig of 3947 bp in length  
 38749 38848: gap of 100 bp  
 38849 48303: contig of 9455 bp in length  
 48304 48403: gap of 100 bp  
 48404 55751: contig of 7348 bp in length  
 55752 55851: gap of 100 bp  
 55852 65044: contig of 9193 bp in length  
 65045 65144: gap of 100 bp  
 65145 74306: contig of 9162 bp in length  
 74307 74406: gap of 100 bp  
 74407 85102: contig of 10696 bp in length  
 85103 85202: gap of 100 bp  
 85203 95862: contig of 10660 bp in length  
 95863 95962: gap of 100 bp  
 95963 111886: contig of 15924 bp in length  
 111887 111986: gap of 100 bp  
 111887 127546: contig of 15560 bp in length  
 127547 127646: gap of 100 bp  
 127647 144833: contig of 17187 bp in length  
 144834 144933: gap of 100 bp  
 144934 162954: contig of 18021 bp in length  
 162955 163054: gap of 100 bp  
 163055 181542: contig of 18488 bp in length.

Location/Qualifiers

1. 181542

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-102H15"

/clone.lib="RPC1-11 Human Male BAC"

1. 1096

/note="assembly-fragment"

1197. 3560

/note="assembly-fragment"

3661. 5983

/note="assembly-fragment"

6084. 8818

/note="assembly-fragment"

8919. 11436

/note="assembly-fragment"

11537. 14541

/note="assembly-fragment"

14642. 18297

/note="assembly-fragment"

18398. 22136

/note="assembly-fragment"

22237. 25665

/note="assembly-fragment"

25766. 30336

/note="assembly-fragment"

30437. 34701

/note="assembly-fragment"

34802. 38748

/note="assembly-fragment"

38849. 48303

/note="assembly-fragment"

48404. 55751

/note="assembly-fragment"

55852. 65044

/note="assembly-fragment"

65145. 74306

/note="assembly-fragment"

74407. 85102

/note="assembly-fragment"

85203. 95862

/note="assembly-fragment"

95963. 111886

/note="assembly-fragment"

111887. 127546

/note="assembly-fragment"

127547. 144833

/note="assembly-fragment"

144834. 144933

/note="assembly-fragment"

144934. 162954

/note="assembly-fragment"

162955. 163054

/note="assembly-fragment"

163055. 181542

/note="assembly-fragment"



BASE COUNT	283 a	254 c	283 g	266 t
ORIGIN				

RESULT	42
LOCUS	AE001152
DEFINITION	AE001152 17078 bp DNA BCT 15-DEC-1997
ACCESSION	AE001152 AE000783
VERSION	AE001152.1 GI:2688387
KEYWORDS	
SOURCE	
ORGANISM	Lyme disease spirochete. Borrelia burgdorferi Bacteria; Spirochaetales; Spirochaetaceae; Borrelia burgdorferi group. 1 (bases 1 to 17078)
REFERENCE	Fraser, C.M., Caajens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lahidra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwin, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D.,
AUTHORS	



TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
2 (bases 1 to 17078)  
Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.  
Direct Submission  
Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. 17078  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
361..1725  
/gene="BB0473"  
361..1725  
/gene="BB0473"  
/note="similar to SP:P54181 PID:1256651 GB:AL009126 percent identity: 23.93. Identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical integral membrane protein"  
/protein\_id="AAC66867.1"  
/db\_xref="GI:2688416"  
/translation="MSTDKSKTREILNGNIXKVFILSPVITNITQAFDLTDM YGKIGAPLSALSLAGPVNFFITAIAMGATGSI LSKCIGESNRSFYAQOLI VLNFLVLEVTICAFEFIDLLDLGVGELKEISRYEYVITFPIPIMFISITVI LNAOGETILSMIVLFANIVNFIIDPILIFSENMGTGAAMATSKLLIVFYELF YRLNKGKILHLDVIDIRISKEIVNLGLPTFGQIMVLSFIFNIVIEISPEFLA AVGLNTIISLSEFLPMAGIGGTISIVGONLGAKKVNEVEYLKFFISALIIITN SIVLENKOTILRFTFNDELEVNANNVLLTIGCGVGLQOVPRFGGLSGSRRTIAM IVIRIRMLIRLIPVVFIFQYFIMENSLGTAFTISNYAIIILIAFTCRIRWAKFILI KKKK"  
1764..1865  
/gene="BB0474"  
1764..1865  
/gene="BB0474"  
/note="hypothetical protein: identified by Glimmer: putative"  
/codon\_start=1  
/transl\_table=1  
/product="B. burgdorferi predicted coding region BB0474"  
/protein\_id="AAC66869.1"  
/db\_xref="GI:2688418"  
/translation="WAYFLINLMFLAKIDFRFFVNLFSVYCLQ"  
complement(1923..2315)  
/gene="BB0475"  
complement(1923..2315)  
/gene="BB0475"  
/note="hypothetical protein: identified by Glimmer: putative"  
/codon\_start=1  
/transl\_table=1  
/product="B. burgdorferi predicted coding region BB0475"  
/protein\_id="AAC66868.1"  
/db\_xref="GI:2688417"  
/translation="YMAQKFKRLKLPILVITSGILIVFMSCKTSTIKENAKETTY LISTIKINOKYEIVNKSQSKNLIITLKNSTEDINANSIAIFEGSKTGLIREKL NGLETFPHLKTINTKRTLTLYIFERK"  
2785..3990  
/gene="BB0476"

CDS  
2785..3990  
/gene="BB0476"  
/note="similar to GB:I23125 PID:349407 SP:P50062 PID:1685369 GB:AE000783 percent identity: 100.00; identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="translation elongation factor TU (tuf)"  
/protein\_id="AAC66866.1"  
/db\_xref="GI:2688415"  
/translation="MKPRIVMAKEVQRTKPHNNGTIGVHGKTTLPAAISYTS KLNDAKAKLEDIDNAPEEKARGITTNARIETANRHRARHDCGADYIKAMIT GAQMDAAILVAADSGAEPTQKCHLLAQMGITKIVELNKDLPDLVEIVEE FLAUEKGFESADPTIKGSAFGAMSNDEPSCVKELESMDNYFDLPERIDKP VLEVEDVESISGRVATGRIERGIIKVGGEVIVGIRKRTYTVGVEMFKILEQ GOAGNVGLLRGVGDKDIERGOVLSAAGTTPPKFKRASICYLCTKEGGRKPFEPG YRPOEFFRTTDTGVVLEGEKEMVPGDNVDIYELLSIAMKNVEFAVREGRTVA SGRILEFE"  
4041..4352  
/gene="BB0477"  
4041..4352  
/gene="BB0477"  
/note="similar to GP:1685370 percent identity: 99.03; identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="ribosomal protein S10 (rpsu)"  
/protein\_id="AAC66865.1"  
/db\_xref="GI:2688414"  
/translation="WIAKDRIKRVLRFSDDYKILDOSASTYKAYOKAKOIKPIPLP TKIKTYLVRSPHNKRSREOFEMKTRRLIDILEPTSLAMDSLMKLELPAGVEVDIK O"  
4388..5008  
/gene="BB0478"  
4388..5008  
/gene="BB0478"  
/note="similar to GP:1685371 percent identity: 99.03; identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="ribosomal protein L3 (rplc)"  
/protein\_id="AAC66864.1"  
/db\_xref="GI:2688413"  
/translation="MLGILGKVKMTQIFQKNGIVPVYVYIEOPNITICKKYVDRG YSALIAGSVDLKSSKSPKIPKIGYKSLKDIPEKRYVELGLDGYDGDIDKVDVRS VKYVDVGTGTRKGGFOGAKMRHNSGGPSSGSKFHHLGDTGATTPARTFPGTKWA GRMGNOQTIONLEVLIDEEKRALLVKAVPAGKGSFVVKRSKK"  
5018..5647  
/gene="BB0479"  
5018..5647  
/gene="BB0479"  
/note="similar to GP:1685372 percent identity: 99.04; identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="ribosomal protein L4 (rpld)"  
/protein\_id="AAC66863.1"  
/db\_xref="GI:2688412"  
/translation="MERKVFESKDKKEIGTINLDDRVNIEISHGSIYNAIKNELSNLR VGTSTKTRSEVRSGSKKPKRGKTGRARVGTGRNVPVIGGIALGPRPDYSYRLPK KVKKIAFQSVSLIRADENSFRKVENNVESGKDKDALIITKNPASPNCVVILLGND DQNIKRAKGNIRDLKILSFDKLRVYDLPYAKNLIALSAVKNLNEFIK"  
5653..5964  
/gene="BB0480"  
5653..5964  
/gene="BB0480"  
/note="similar to GP:1685373 percent identity: 100.00; identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="ribosomal protein L23 (rplw)"  
/protein\_id="AAC66862.1"  
/db\_xref="GI:2688411"



Brandon,R.C., Rogers,Y.H., Blazey,R.G., Champe,M., Pfeiffer,B.D.,  
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
 Miklos,G.L., Abill,J.F., Agbayan,A., An,H.J.,  
 Andrews-Pfannkoch,C., Baldwin,D., Bailley,R.M., Basu,A.,  
 Baxendale,J.R., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
 Benos,P.V., Berman,B.P., Bhargava,D., Bolshakov,S., Borlova,D.,  
 Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burris,C.,  
 Busam,D.A., Butler,H., Dahlu,E., Center,A., Chandra,I.,  
 Cherry,J.M., Cawley,S., Dahle,C., Davenport,L.B., Davies,P., de  
 Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
 Dodson,K., Doup,I.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
 Dunn,P., Dublin,K.J., Evangelista,C.C., Ferraz,C., Fertlra,S.,  
 Fleischman,W., Foster,C., Gabriellian,A.E., Garg,N.S.,  
 Gelbart,W.M., Glasser,K., Glodex,A., Gong,F., Gorrell,J.H., Gu,Z.,  
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,  
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,  
 Wei,M.H., Idegum,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,  
 Kennison,J.A., Ketchum,K.A., Kimmel,P.E., Kodira,C.D., Kraft,C.,  
 Kravitz,S., Kulp,D., Lai,Z., Lasco,P., Lei,Y., Levitsky,A.A.,  
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,  
 McLeod,M.P., McPherson,D., Merkulov,G., Mishina,N.V., Mobry,C.,  
 Morris,J., Mosthefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,  
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
 Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,  
 Pollard,J., Puri,V., Reese,M.G., Reinett,K., Remington,K.,  
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Klamos,I.,  
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,  
 Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,  
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,  
 Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T.,  
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
 Zaveri,U.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
 The genome sequence of *Drosophila melanogaster*  
 Science 287 (5461), 2185-2195 (2000)

TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

FEATURES  
 source  
 Location/Qualifiers

1. 49759  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="2L"  
 /map="38C1-38C2"  
 <8107..>8511  
 /gene="CG17570"  
 /product="CT34146"  
 /db\_xref="FLYBASE:FBgn0017570"  
 /db\_xref="FLYBASE:FBgn0032948"  
 /evidence=not\_experimental  
 <8107..>8511  
 /gene="CG17570"  
 /map="39E1-39E1"  
 /db\_xref="FLYBASE:FBgn0017570"  
 /db\_xref="FLYBASE:FBgn0032948"  
 /evidence=not\_experimental  
 8107..8511  
 /gene="CG17570"  
 /note="CG17570 gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBgn0017570"  
 /db\_xref="FLYBASE:FBgn0032948"  
 /evidence=not\_experimental  
 /protein\_id="AA045446.1"  
 /db\_xref="GI:7299775"  
 /translation="MNIYALKTSASYPPLPGHKGDEGNITLADLQOISANFPH  
 AINSYKTPSEKRTVFIAKKWHIOMIMPEFLPRKQIRLRGKHARVYLEHYFOOE  
 LPELVIEKTIYVNGFLSRDKRNLLAQQPATRP"  
 <15476..16286

mRNA



gene

CDs

BASE COUNT 14487 a 10540 c 10241 g 14491 t

ORIGIN

Query Match 23.4%; Score 28.8; DB 3; Length 49759;  
Best Local Similarity 58.0%; Pred. No. 30;  
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 34 tgatgactaataagcgcaatgcagggctgcactgctgtaagttcacaatgtaagtg 93  
Db 12852 TGGGGAACGCTGCTACAAAGCGGCTGTTAGTTAGTCATCACTGAGCAG 12793  
QY 94 ataactatctcaattatccacaa 121  
Db 12792 ACAATTAAGCTGCTTGCATCCATTCGA 12765

RESULT 44  
AC004735/c  
LOCUS  
DEFINITION  
AC004735 74534 bp DNA INV 31-OCT-2000  
Drosophila melanogaster. chromosome 2L, region 38A7-38C9, P1 clone  
DS04178, complete sequence.  
AC004735 AC004119 AC000718 AC000720 AC000721 L81497  
AC001816 L81498 AC000722 AC000723 AC001208 L81496 AC000724  
AC000725 AC000726 AC000727  
AC004735.1 GI:3892961  
HTG.  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS  
Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshirefi, A.R., Moshirefi, M., Nixon, K., Pacle, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Punch, E., Sequeira, A., Sethi, H., Shit, E.,  
Svirskas, R.R., Twomey, B., Wan, K.H., Weinburg, T., Zhang, R.,  
Zliern, L.L. and Rubin, G.M.  
Sequencing of Drosophila chromosome 2L, region 38C  
Unpublished  
2 (bases 1 to 74534)  
Celisner, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,  
Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,  
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,  
Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,  
Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshirefi, A.R.,  
Moshirefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B., Punch, E.,

SNL, E., Twomey, B., Wan, K.H., White, J., Yee, A., Zhang, R.,  
Zliern, L.L. and Kimmel, B.E.  
Direct Submission  
Submitted (22-MAY-1998) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Nov 19, 1998 this sequence version replaced gi:3766072  
gi:2829852.  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).  
Pl library locations: 44-50.  
FEATURES  
source  
1. 74534  
/organism="Drosophila melanogaster"  
/strain="y2; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="38A7-38C9"  
/clone="P1\_DS04178 (D78)"  
/clone\_lib="P1 library, partial Sau3a in pAd10sacBII"  
BASE COUNT 22626 a 15319 c 15152 g 21437 t

ORIGIN

Query Match 23.4%; Score 28.8; DB 3; Length 74534;  
Best Local Similarity 58.0%; Pred. No. 30;  
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 34 tgatgactaataagcgcaatgcagggctgcactgctgtaagttcacaatgtaagtg 93  
Db 48537 TGGGGAACGCTGCTACAAAGCGGCTGTTAGTTAGTCATCACTGAGCAG 48478  
QY 94 ataactatctcaattatccacaa 121  
Db 48477 ACAATTAAGCTGCTTGCATCCATTCGA 48450

RESULT 45  
AL136093  
LOCUS  
DEFINITION  
AL136093 85195 bp DNA PRI 30-SEP-2000  
Human DNA sequence from clone RP3-472A9 on chromosome 6q13-14.3,  
complete sequence.  
AC004735 AC000722 AC000723 AC001208 L81496 AC000724  
AC000725 AC000726 AC000727  
AC004735.1 GI:10443356  
HTG.  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS  
Matthews, L.  
Direct Submission  
Submitted (30-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquerry@sanger.ac.uk](mailto:humquerry@sanger.ac.uk) Clone  
requests: [clonerequests@sanger.ac.uk](mailto:clonerequests@sanger.ac.uk)  
On Oct 1, 2000 this sequence version replaced gi:19931221.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unSURE'  
feature key.







**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:34:11 ; Search time 100.26 Seconds

(without alignments)  
1051.775 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 cagtgatgttaatgtgaga.....tcaattatccacacacaca 123

Scoring table: IDENTITY\_NUC

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_1101.\*  
1: /SIDS2/gcgcdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDS2/gcgcdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDS2/gcgcdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDS2/gcgcdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDS2/gcgcdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDS2/gcgcdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDS2/gcgcdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDS2/gcgcdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDS2/gcgcdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDS2/gcgcdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDS2/gcgcdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDS2/gcgcdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDS2/gcgcdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDS2/gcgcdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDS2/gcgcdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDS2/gcgcdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDS2/gcgcdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDS2/gcgcdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDS2/gcgcdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDS2/gcgcdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDS2/gcgcdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDS2/gcgcdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.8	42.1	1422	21	AAc44809
2	49.6	40.3	606	21	AAc42911
3	49.6	40.3	913	21	AAc46260
4	49.6	40.3	1204	21	AAc33604
5	49.6	40.3	1272	21	AAc33124
6	47	38.2	1104	21	AAc46355
7	47	38.2	1145	21	AAc45592
8	47	38.2	1148	21	AAc32774
9	34.4	28.0	1082138	21	AAc22305
10	28.8	23.4	910715	20	AAx20248
11	28.4	23.1	582	21	AAc49264

C	12	26.4	21.5	1474	21	AAc45107	Arabidopsis thalia
C	13	26.4	21.5	2347	21	AAz46167	CDNA sequence enco
C	14	26	21.1	2620	22	AAH15633	Human CDNA sequenc
C	15	25.6	20.8	634	20	AAH20838	Poly nucleotide seq
C	16	25.6	20.8	7147	19	AAV52156	Streptococcus pneu
C	17	25.6	20.8	19932	17	AAV46159	CagI locus, Helic
C	18	25.4	20.7	591	21	AAZ88746	A. dichotoma kabut
C	19	25.4	20.7	2333	22	AAH17997	Human CDNA sequenc
C	20	25.4	20.7	3300	21	AAH92497	Haemophilus Influe
C	21	25.2	20.5	649	22	AAH83942	Callitrix jaccus
C	22	25.2	20.5	3684	22	AAI60262	Human polynucleoti
C	23	25.2	20.5	12412	17	AAH33537	BCG deletion regio
C	24	25	20.3	933	22	AAH32402	Human olfactory re
C	25	25	20.3	2284	20	AAH04441	S. frugiperda immu
C	26	25	20.3	49998	20	AAH23518	Human kidney amino
C	27	24.8	20.2	528	22	AAH13128	Human CDNA clone (
C	28	24.8	20.2	1448	21	AAH39501	Arabidopsis thalia
C	29	24.8	20.2	1782	22	AAH23414	Human SEC5 DNA seq
C	30	24.8	20.2	1970	22	AAH16592	Human CDNA sequenc
C	31	24.8	20.2	2064	16	AAH80535	Paralichthys oliva
C	32	24.8	20.2	2064	16	AAH80533	Paralichthys oliva
C	33	24.8	20.2	2148	14	AAH46026	Transglutaminase (
C	34	24.8	20.2	2148	14	AAH54878	Transglutaminase (
C	35	24.8	20.2	2148	16	AAH80536	Paralichthys oliva
C	36	24.8	20.2	2148	16	AAH80534	Paralichthys oliva
C	37	24.6	20.0	452	21	AAH59911	Human secreted pro
C	38	24.6	20.0	625	22	AAH26556	DNA encoding human
C	39	24.6	20.0	1864	8	AAH71405	Sequence of ANS-1
C	40	24.6	20.0	2946	22	AAH62796	Shrimp white spot
C	41	24.6	20.0	3787	22	AAH81775	Human differential
C	42	24.6	20.0	68940	20	AAH57351	Human chromosome 6
C	43	24.6	20.0	109519	22	AAH58693	Micromonospora DNA
C	44	24.4	19.8	294	21	AAH75339	Human ORF ORF894
C	45	24.4	19.8	300	21	AAH01147	Human colon cancer

# ALIGNMENTS

RESULT	1
ID	AAc44809 standard; DNA: 1422 BP.
AC	AAc44809;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Zea mays DNA fragment SEQ ID NO: 44214.
XX	
XX	Hybridisation assay; genetic mapping; gene expression control;
KW	Protein identification; signal transduction pathway; metabolic;
KW	pathway; promoter; termination sequence; corn; ss.
XX	
OS	Zea mays subsp. mays.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.



```
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0151363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158365.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
```



PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 42.1%; Score 51.8; DB 21; Length 1422;  
Best Local Similarity 83.1%; Pred. No. 6.2e-09;  
Matches 59; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 4 gttatgtaatgtagatgcttaccctgtatgacctatgagcgaatgcagagctg 63  
Db 783 gttatgcaagtgtgagatgcttaccacccagacacccatcgttgcagtgtgaagctg 842

OY 64 cactgactggt 74  
Db 843 caagactggt 853

## RESULT 2

AAAC42911  
ID AAAC42911 standard; DNA; 606 BP.

AC AAAC42911;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 37322.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.



PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160985.
PR	22-OCT-1999;	99US-0160987.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161355.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match		
Best Local Similarity	40.3%;	Score 49.6; DB 21; Length 606;
Matches	58; Conservative	80.6%; Pred. No. 2,8e-08;
		0; Mismatches 14; Indels 0; Caps
OY	4	gatatgttaatgtgagatgccttaaacccgtgatgactaatggcgaatgcgaggtcg 63
Db	342	gactcgaatgtgaatgccttaacatccagacgatctcatgtgcagttgcaaggtg 401
OY	64	cactgactgcta 75
Db	402	caagactcgta 413
RESULT 3		
AAC46260	ID AAC46260 standard; DNA; 913 BP.	
XX	AAC46260;	
XX	18-OCT-2000 (first entry)	
DE	Arabidopsis thaliana DNA fragment SEQ ID.NO: 49491.	
XX		
KW	Hybridisation assay; genetic mapping; gene expression control;	
KM	protein identification; signal transduction pathway;	
KW	metabolic pathway; promoter; termination sequence; ss.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.



PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134258.  
PR 14-MAY-1999; 99US-0134259.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155386.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158399.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.



```
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 40.3%; Score 49.6; DB 21; Length 913;
Best Local Similarity 80.6%; Pred. No. 3.3e-08;
Matches 58; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 4 gttatgttaattgtgagtgtcttacaacctgtgacctaatgctgaatgcgaatgcgaagctg 63
Db 344 gttatgttaattgtgagtgtcttacaacctgtgacctaatgctgaatgcgaatgcgaagctg 403
Qy 64 cactgactgtgta 75
Db 404 caaagactgtgta 415

RESULT 4
AAC33604
ID AAC33604 standard; DNA; 1204 BP.
XX AAC33604;
AC 17-OCT-2000 (first entry)
XX
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 3658.
DE
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
XX
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131448.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
```

```
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
```







PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160776.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.



[illegible]

PR	16-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-0134941
PR	20-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135533
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136392
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137722
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139463
PR	17-JUN-1999	9905-0139493
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0141887
PR	22-JUN-1999	9905-0139869
PR	23-JUN-1999	9905-0140353
PR	24-JUN-1999	9905-0140695
PR	28-JUN-1999	9905-0140823
PR	29-JUN-1999	9905-0140991
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0142547
PR	01-JUL-1999	9905-0142155
PR	02-JUL-1999	9905-0142390
PR	06-JUL-1999	9905-0142803
PR	08-JUL-1999	9905-0142820
PR	09-JUL-1999	9905-0142920
PR	12-JUL-1999	9905-0142977
PR	13-JUL-1999	9905-0143542
PR	14-JUL-1999	9905-0143624
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144357
PR	20-JUL-1999	9905-0144684
PR	21-JUL-1999	9905-0144614
PR	21-JUL-1999	9905-0145086
PR	21-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145152
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145224
PR	26-JUL-1999	9905-0145276
PR	27-JUL-1999	9905-0145513
PR	27-JUL-1999	9905-0145518



PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-015659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 38.2%; Score 47; DB 21; Length 1104;  
Best Local Similarity No. 98; Pred. No. 3e-07;  
Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
  
Qy 4 gttctgtaaatgtgagatgcttaacacctgatacctaattgycgcgaatgcgagagctg 63  
Db 837 gtcttgcgaagtgtgagatgcttaacacctgatacctaattgctggtgcaatgcgagagctg 896  
  
Qy 64 cactgactggt 74  
Db 897 ttctgagtgt 907  
  
RESULT 7  
AAC45592  
ID AAC45592 standard; DNA; 1145 BP.  
XX  
AC AAC45592;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 47055.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 08-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0129845.  
PR 21-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.



PR 21-MAY-1999;	990S-0135353.
PR 24-MAY-1999;	990S-0135629.
PR 25-MAY-1999;	990S-0136021.
PR 27-MAY-1999;	990S-0136392.
PR 28-MAY-1999;	990S-0136782.
PR 01-JUN-1999;	990S-0137222.
PR 03-JUN-1999;	990S-0137528.
PR 04-JUN-1999;	990S-0137502.
PR 07-JUN-1999;	990S-0137724.
PR 08-JUN-1999;	990S-0138094.
PR 10-JUN-1999;	990S-0138540.
PR 10-JUN-1999;	990S-0138847.
PR 14-JUN-1999;	990S-0139119.
PR 16-JUN-1999;	990S-0139452.
PR 16-JUN-1999;	990S-0139453.
PR 17-JUN-1999;	990S-0139492.
PR 18-JUN-1999;	990S-0139454.
PR 18-JUN-1999;	990S-0139455.
PR 18-JUN-1999;	990S-0139456.
PR 18-JUN-1999;	990S-0139457.
PR 18-JUN-1999;	990S-0139458.
PR 18-JUN-1999;	990S-0139459.
PR 18-JUN-1999;	990S-0139460.
PR 18-JUN-1999;	990S-0139461.
PR 18-JUN-1999;	990S-0139462.
PR 18-JUN-1999;	990S-0139463.
PR 18-JUN-1999;	990S-0139750.
PR 18-JUN-1999;	990S-0139763.
PR 21-JUN-1999;	990S-0139817.
PR 22-JUN-1999;	990S-0139899.
PR 23-JUN-1999;	990S-0140353.
PR 23-JUN-1999;	990S-0140354.
PR 24-JUN-1999;	990S-0140695.
PR 28-JUN-1999;	990S-0140823.
PR 29-JUN-1999;	990S-0140991.
PR 30-JUN-1999;	990S-0141287.
PR 01-JUL-1999;	990S-0141842.
PR 01-JUL-1999;	990S-0142154.
PR 02-JUL-1999;	990S-0142055.
PR 06-JUL-1999;	990S-0142390.
PR 08-JUL-1999;	990S-0142803.
PR 09-JUL-1999;	990S-0142920.
PR 12-JUL-1999;	990S-0142977.
PR 13-JUL-1999;	990S-0143542.
PR 14-JUL-1999;	990S-0143624.
PR 15-JUL-1999;	990S-0144005.
PR 16-JUL-1999;	990S-0144085.
PR 16-JUL-1999;	990S-0144086.
PR 19-JUL-1999;	990S-0144325.
PR 19-JUL-1999;	990S-0144331.
PR 19-JUL-1999;	990S-0144332.
PR 19-JUL-1999;	990S-0144333.
PR 19-JUL-1999;	990S-0144334.
PR 19-JUL-1999;	990S-0144335.
PR 20-JUL-1999;	990S-0144352.
PR 20-JUL-1999;	990S-0144632.
PR 20-JUL-1999;	990S-0144884.
PR 21-JUL-1999;	990S-0144814.
PR 21-JUL-1999;	990S-0145086.
PR 21-JUL-1999;	990S-0145088.
PR 22-JUL-1999;	990S-0145085.
PR 22-JUL-1999;	990S-0145087.
PR 22-JUL-1999;	990S-0145089.
PR 22-JUL-1999;	990S-0145192.
PR 23-JUL-1999;	990S-0145145.
PR 23-JUL-1999;	990S-0145224.
PR 26-JUL-1999;	990S-0145276.
PR 27-JUL-1999;	990S-0145913.
PR 27-JUL-1999;	990S-0145918.
PR 27-JUL-1999;	990S-0145919.
PR 28-JUL-1999;	990S-0145951.
PR 02-AUG-1999;	990S-0146386.
PR 02-AUG-1999;	990S-0146388.
PR 02-AUG-1999;	990S-0146389.
PR 03-AUG-1999;	990S-0147038.
PR 04-AUG-1999;	990S-0147204.
PR 04-AUG-1999;	990S-0147302.
PR 05-AUG-1999;	990S-0147192.
PR 05-AUG-1999;	990S-0147260.
PR 06-AUG-1999;	990S-0147303.
PR 06-AUG-1999;	990S-0147416.
PR 09-AUG-1999;	990S-0147493.
PR 09-AUG-1999;	990S-0147935.
PR 10-AUG-1999;	990S-0148171.
PR 11-AUG-1999;	990S-0148319.
PR 12-AUG-1999;	990S-0148341.
PR 13-AUG-1999;	990S-0148345.
PR 13-AUG-1999;	990S-0148684.
PR 16-AUG-1999;	990S-0149368.
PR 17-AUG-1999;	990S-0149175.
PR 18-AUG-1999;	990S-0149426.
PR 20-AUG-1999;	990S-0149722.
PR 20-AUG-1999;	990S-0149723.
PR 20-AUG-1999;	990S-0149929.
PR 23-AUG-1999;	990S-0149902.
PR 23-AUG-1999;	990S-0149930.
PR 25-AUG-1999;	990S-0150566.
PR 26-AUG-1999;	990S-0150884.
PR 27-AUG-1999;	990S-0151065.
PR 27-AUG-1999;	990S-0151066.
PR 27-AUG-1999;	990S-0151080.
PR 30-AUG-1999;	990S-0151303.
PR 31-AUG-1999;	990S-0151438.
PR 01-SEP-1999;	990S-0151930.
PR 07-SEP-1999;	990S-0152363.
PR 10-SEP-1999;	990S-0153070.
PR 13-SEP-1999;	990S-0153758.
PR 15-SEP-1999;	990S-0154018.
PR 16-SEP-1999;	990S-0154039.
PR 20-SEP-1999;	990S-0154779.
PR 22-SEP-1999;	990S-0155139.
PR 23-SEP-1999;	990S-0155486.
PR 24-SEP-1999;	990S-0155659.
PR 28-SEP-1999;	990S-0156458.
PR 29-SEP-1999;	990S-0156596.
PR 04-OCT-1999;	990S-0157117.
PR 05-OCT-1999;	990S-0157753.
PR 06-OCT-1999;	990S-0157865.
PR 07-OCT-1999;	990S-0158029.
PR 08-OCT-1999;	990S-0158232.
PR 12-OCT-1999;	990S-0158369.
PR 13-OCT-1999;	990S-0159293.
PR 13-OCT-1999;	990S-0159294.
PR 13-OCT-1999;	990S-0159295.
PR 14-OCT-1999;	990S-0159329.
PR 14-OCT-1999;	990S-0159330.
PR 14-OCT-1999;	990S-0159331.
PR 14-OCT-1999;	990S-0159337.
PR 14-OCT-1999;	990S-0159638.
PR 18-OCT-1999;	990S-0159584.
PR 21-OCT-1999;	990S-0160741.
PR 21-OCT-1999;	990S-0160767.
PR 21-OCT-1999;	990S-0160768.
PR 21-OCT-1999;	990S-0160770.
PR 21-OCT-1999;	990S-0160814.
PR 21-OCT-1999;	990S-0160815.
PR 22-OCT-1999;	990S-0160980.
PR 22-OCT-1999;	990S-0160981.
PR 22-OCT-1999;	990S-0160989.
PR 25-OCT-1999;	990S-0161404.
PR 25-OCT-1999;	990S-0161405.
PR 25-OCT-1999;	990S-0161406.
PR 26-OCT-1999;	990S-0161359.
PR 26-OCT-1999;	990S-0161360.
PR 26-OCT-1999;	990S-0161361.



PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
Query Match 38.28; Score 47; DB 21; Length 1145;  
Best Local Similarity 78.98; Pred. No. 3,1e-07;  
Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
OY 4 gttatgtaaatgtatgagcttacaacctgatgcacctaatacgcgagagctg 63  
Db 566 gtctctgcaagtgtagagatgccgtataaccacgatgacttgatgctgaatgcgagagctg 625  
OY 64 caactgactggt 74  
Db 626 ttctcgactggt 636  
RESULT 8  
AAC32774  
ID AAC32774 standard; DNA; 1148 BP.  
XX AAC32774;  
AC AAC32774;  
XX 17-OCT-2000 (first entry)  
DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 602.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
KM metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138040.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.



PR	04-AUG-1999	9905-01473202
PR	04-AUG-1999	9905-01473201
PR	05-AUG-1999	9905-01471192
PR	05-AUG-1999	9905-01472760
PR	06-AUG-1999	9905-01473703
PR	06-AUG-1999	9905-01474761
PR	06-AUG-1999	9905-01474763
PR	09-AUG-1999	9905-01479325
PR	09-AUG-1999	9905-01479325
PR	10-AUG-1999	9905-01481171
PR	11-AUG-1999	9905-01483119
PR	12-AUG-1999	9905-01485341
PR	13-AUG-1999	9905-01486864
PR	16-AUG-1999	9905-01493365
PR	17-AUG-1999	9905-01491375
PR	18-AUG-1999	9905-01494226
PR	20-AUG-1999	9905-01497422
PR	20-AUG-1999	9905-01497233
PR	20-AUG-1999	9905-01499329
PR	23-AUG-1999	9905-01499302
PR	23-AUG-1999	9905-01499303
PR	25-AUG-1999	9905-01505666
PR	26-AUG-1999	9905-01508844
PR	27-AUG-1999	9905-01510655
PR	27-AUG-1999	9905-01510666
PR	27-AUG-1999	9905-01511080
PR	30-AUG-1999	9905-01511033
PR	31-AUG-1999	9905-01514338
PR	01-SEP-1999	9905-01519330
PR	07-SEP-1999	9905-01523263
PR	10-SEP-1999	9905-01530707
PR	13-SEP-1999	9905-01537358
PR	15-SEP-1999	9905-01540319
PR	16-SEP-1999	9905-01540339
PR	20-SEP-1999	9905-01551339
PR	23-SEP-1999	9905-01554866
PR	24-SEP-1999	9905-01556558
PR	28-SEP-1999	9905-01565649
PR	04-OCT-1999	9905-01569167
PR	05-OCT-1999	9905-01577513
PR	06-OCT-1999	9905-01578655
PR	07-OCT-1999	9905-01580289
PR	08-OCT-1999	9905-01583329
PR	12-OCT-1999	9905-01583639
PR	13-OCT-1999	9905-01592933
PR	13-OCT-1999	9905-01592934
PR	13-OCT-1999	9905-01595925
PR	14-OCT-1999	9905-01593329
PR	14-OCT-1999	9905-01593329
PR	14-OCT-1999	9905-01595931
PR	14-OCT-1999	9905-01596337
PR	14-OCT-1999	9905-01596338
PR	18-OCT-1999	9905-01595884
PR	21-OCT-1999	9905-01607411
PR	21-OCT-1999	9905-01609881
PR	21-OCT-1999	9905-01607657
PR	21-OCT-1999	9905-01607658
PR	25-OCT-1999	9905-01610484
PR	25-OCT-1999	9905-01614004
PR	25-OCT-1999	9905-01614005
PR	25-OCT-1999	9905-01614006
PR	26-OCT-1999	9905-01613569
PR	26-OCT-1999	9905-01613661
PR	26-OCT-1999	9905-01613661
PR	28-OCT-1999	9905-01619320
PR	28-OCT-1999	9905-01619320
PR	28-OCT-1999	9905-01619393

PR 29-OCT-1999; 99US-0162142.

Query Match

38.28; Score 47; DB 21; Length 1148;

Matches	56;	Conservative	0;	Mismatches	15;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

QY

4 gtaattgaatgtgagatgccttacaacccgtgatgacctaatg'gcgaatg'cgaggcctg 63

Dh

568 gttctgcaagtgtgagatgccgtatataccagatgacttgatgtgcacatgcgaggaagtg 627

QY

64 cactgactgt 74

Dh

628 ttctgagtggt 638

## RESULT 9

ID AAF22305 standard; DNA; 1082138 BP.

AC AAF22305-

DT 20-MAR-2001 (first entry)

DE *Arabidopsis thaliana* chromosome 4 centromere.

KW Centromere; michromosome; vector; ds.

OS *Arabidopsis thaliana*.

PN WO200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000WO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR	18-MAY-1999;	99US-0134770.
PR	13-SEP-1999	00US-0152594

PR 17-SEP-1999; 99US-0154603.

PA (UYCH-) UNIV CHICAGO.

Preuss D, Copenhaver G, Keith K;

DR WPI; 2000-587529/55

PT Recombinant DNA construct comprising a plant centromere, useful for producing stably inherited microsomas which can serve as vectors for

PT the construction of transgenic plant and animal cells

PS Claim 68; Page 977-1388; 1449pp; English.  
XY

cc The present invention relates to a recombinant DNA construct of a plant  
cc (*Arabidopsis thaliana*) centromere. The constructs are useful for

cc producing stably inherited mitochondria which can serve as vectors for

cc selected proteins such as hormones, enzymes, interleukins, clotting factors, cytokines, antibodies, and growth factors.

Sequence 1082138 BP: 348775 A: 194404 C: 195515 G: 343444 T: 0 other:

Very Match	28.08;	Score 34.4;	Dd 21;	Beunglin 1002130
Best Local Similarity	78.8%;	Pred. No. 0.11;		

MACHINES	41,	CONSERVATIVE	V,	MISMAKINGS	11,	INDEXES	V,	CUPS	V,
----------	-----	--------------	----	------------	-----	---------	----	------	----

27

[illegible]

RESULT 10

100







PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144881.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 23-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 23.1%; Score 28.4; DB 21; Length 582;  
Best Local Similarity 76.1%; Pred. No. 1.2; Indels 0; Gaps 0;  
Matches 35; Conservative 0; Mismatches 11;

Oy 4 gtaattgaatgagatgccttaacaacctgagctaatgagc 49  
||| | ||||| |||| | ||| | |||| | |||| |  
Db 327 gtactacaatgtgaatgcgtatactacagatgactgagtgg 372

RESULT 12  
AAC45107/c  
ID AAC45107 standard; DNA; 1474 BP.  
XX  
AC AAC45107;  
XX  
18-OCT-2000 (first entry)  
XX  
Arabidopsis thaliana DNA fragment SEQ ID NO: 45325.  
DE  
XX  
Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX



OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145066.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145102.  
PR 22-JUL-1999; 99US-0145109.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.



PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.  
PR 20-SEP-1999; 990S-0154779.  
PR 22-SEP-1999; 990S-0155139.  
PR 23-SEP-1999; 990S-0155486.  
PR 24-SEP-1999; 990S-0155659.  
PR 28-SEP-1999; 990S-0156458.  
PR 29-SEP-1999; 990S-0156596.  
PR 04-OCT-1999; 990S-0157117.  
PR 05-OCT-1999; 990S-0157753.  
PR 06-OCT-1999; 990S-0157865.  
PR 07-OCT-1999; 990S-0158029.  
PR 08-OCT-1999; 990S-0158232.  
PR 12-OCT-1999; 990S-0158369.  
PR 13-OCT-1999; 990S-0159293.  
PR 13-OCT-1999; 990S-0159294.  
PR 13-OCT-1999; 990S-0159295.  
PR 14-OCT-1999; 990S-0159329.  
PR 14-OCT-1999; 990S-0159330.  
PR 14-OCT-1999; 990S-0159331.  
PR 14-OCT-1999; 990S-0159637.  
PR 14-OCT-1999; 990S-0159638.  
PR 18-OCT-1999; 990S-0159584.  
PR 21-OCT-1999; 990S-0160741.  
PR 21-OCT-1999; 990S-0160767.  
PR 21-OCT-1999; 990S-0160768.  
PR 21-OCT-1999; 990S-0160770.  
PR 21-OCT-1999; 990S-0160815.  
PR 21-OCT-1999; 990S-0160814.  
PR 22-OCT-1999; 990S-0160980.  
PR 22-OCT-1999; 990S-0160981.  
PR 22-OCT-1999; 990S-0160989.  
PR 25-OCT-1999; 990S-0161404.  
PR 25-OCT-1999; 990S-0161405.  
PR 25-OCT-1999; 990S-0161406.  
PR 26-OCT-1999; 990S-0161359.  
PR 26-OCT-1999; 990S-0161360.  
PR 26-OCT-1999; 990S-0161361.  
PR 28-OCT-1999; 990S-0161920.  
PR 28-OCT-1999; 990S-0161992.  
PR 28-OCT-1999; 990S-0161993.  
PR 29-OCT-1999; 990S-0162142.  
Query Match 21.5%; Score 26.4; DB 21; Length 1474;  
Best Local Similarity 61.8%; Pred. No. 8.8;  
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 55 cgaaggctgcaactgactgtaagttcacatgtaagtgatactatcttcaattatc 114  
DB 937 CGATGGCTCTCTGACTGAGAGGTCTCTTGCCTGTGTGTTCCCAACAATTTCTTC 878  
QY 115 cacacaac 122  
DB 877 ACCACCAC 870

RESULT 13  
AAZ46167  
ID AAZ46167 standard; cDNA; 2347 BP.

XX AAZ46167;  
XX 16-MAY-2000 (first entry)  
XX  
DE cDNA sequence encoding a human phosphorylation effector PHSP-30.  
XX  
XX Human; phosphorylation effector; PHSP; proliferative disorder;  
KW immune disorder; neuronal disorder; ss.  
XX  
OS Homo sapiens.  
XX  
XX Key  
FH CDS Location/Qualifiers  
62..1132

FT FT /\*tag= a  
FT FT /product= "phosphorylation effector"  
XX  
PN WO200006728-A2.  
PD 10-FEB-2000.  
PE 28-JUL-1999; 99WO-US17132.  
PF  
XX 28-JUL-1998; 98US-0123494.  
PR 14-SEP-1998; 98US-0152814.  
PR 14-OCT-1998; 98US-0173482.  
PR 03-NOV-1998; 98US-0106889.  
PR 19-NOV-1998; 98US-0109093.  
PR 22-DEC-1998; 98US-0113796.  
PR 12-JAN-1999; 99US-0173482.  
PR 12-JAN-1999; 99US-0229005.  
XX  
XX (INCYTE PHARM INC.  
XX  
XX Hillman JL, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;  
PI Patterson C, Bandman O, Au-Young J, Gorgone GA, Yue H, Azimzai Y;  
PI Reddy R, Lu DM, Smith LL;  
XX WPI: 2000-183125/16.  
XX P-PSDB; AAY68798.  
XX  
XX New human phosphorylation effectors useful for the diagnosis, treatment  
PT and prevention of proliferative, immune and neuronal disorders -  
XX  
XX Claim 9; Page 141-142; 142pp; English.

XX  
XX AAZ46138-246168 encode human phosphorylation effectors (PHSP).  
CC designated PHSP1-PHSP11 (the protein sequence for PHSP28 is not given  
CC in the specification). The sequences were isolated from cDNA libraries  
CC prepared from various human tissues. The PHSP proteins are useful for  
CC the diagnosis, treatment and prevention of proliferative disorders,  
CC immune disorders and neuronal disorders. The PHSP proteins form  
CC pharmaceutical compositions which useful for treating or preventing  
CC disorders associated with decreased PHSP expression/activity. PHSP  
CC antagonists are useful for treating or preventing disorders associated  
CC with increased PHSP expression/activity.  
XX

Sequence 2347 BP; 745 A; 442 C; 454 G; 706 T; 0 other;

Query Match 21.5%; Score 26.4; DB 21; Length 2347;  
Best Local Similarity 57.1%; Pred. No. 10;  
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 29 aaccctgactgactgactgcaatgagagctgcactgactgtaagttcacatgta 88  
DB 34 agccacaatcaccgcgtgtgtagccagccatggtgagtgagtgagaaagcgaaga 93  
QY 89 agtgatactatcttcaattat 112  
DB 94 agatgaaataatcacttcaagctat 117

RESULT 14  
AAH15633/C  
ID AAH15633 standard; cDNA; 2620 BP.

XX AAH15633;  
XX 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:13969.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
KW Homo sapiens.  
XX  
XX  
XX



PN EP1074617-A2.  
PD 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 13969; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2620 BP; 788 A; 581 C; 456 G; 795 T; 0 other;

Query Match 21.1%; Score 26; DB 22; Length 2620;  
Best Local Similarity 52.8%; Pred. No. 15;  
Matches 56; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tattgtaaatgtgagatgacctacacccgtatgacctaatgagcgcaatgagggctgc 64  
DB 1339 TAGAGCACAGAGAGAAAGCTCTTAACCGCTTATGGAGCACTTACACGAGAGATGCG 1280  
OY 65 acggaactgtaagtctacacatgtaagtgatgactatctcaatt 110  
DB 1279 ACAGAGTGAAGATCCATTTTAAAGCTGATGATGATTAATATTTT 1234

RESULT 15  
AAV52156/C  
ID AAV52156 standard; DNA: 634 BP.  
XX  
AC AAV52156;  
XX  
DT 05-MAY-1999 (first entry)  
XX

DE Polynucleotide sequence from the genome of *Treponema pallidum*.  
XX  
XX *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;  
KW enzyme production; ds.  
XX  
XX *Treponema pallidum*.  
OS  
XX  
PN WO9859034-A2.  
XX  
PD 30-DEC-1998.  
XX  
PF 23-JUN-1998; 98WO-US13041.  
XX  
PR 24-JUN-1997; 97US-0050667.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Fraser CM;  
XX  
DR WPI; 1999-081273/07.  
XX  
XX New isolated *Treponema pallidum* nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of *T. pallidum* infections, particularly syphilis  
PT  
XX  
XX Claim 1; Page 929; 1150pp; English.  
PS  
XX  
XX AAX20500-21243 represent polynucleotide sequences from the genome of  
CC *Treponema pallidum*. The sequences can be used for detection,  
CC diagnosis, characterisation, prevention and therapy for *T. pallidum*  
CC infections, particularly syphilis. They can also be used for detecting  
CC diseases related to *Borrelia* infections in animals, and for the  
CC production of biosynthetic products such as enzymes.  
XX  
SQ Sequence 634 BP; 202 A; 127 C; 148 G; 155 T; 2 other;

Query Match 20.8%; Score 25.6; DB 20; Length 634;  
Best Local Similarity 62.5%; Pred. No. 13;  
Matches 40; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 49 gaaatgcagagggctgactgactgtaagtacacatgtaagtgatgactatctcaat 108  
DB 348 gctatgagaattctgacttcttactcaatttacttggaaacccaactgctcttaaaa 407  
OY 109 ttat 112  
DB 408 atat 411

RESULT 16  
AAV52156/C  
ID AAV52156 standard; DNA: 7147 BP.  
XX  
AC AAV52156;  
XX  
DT 23-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:23.  
XX  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9818931-A2.  
XX  
PD 07-MAY-1998.  
XX  
PF 30-OCT-1997; 97WO-US19588.  
XX  
PR 31-OCT-1996; 96US-0029960.  
XX



PA (HUMA-) HUMAN GENOME SCI INC.  
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M.  
PI Kunsch CA, Rosen CA.  
XX WPI: 1998-272225/24.  
DR  
XX  
XX Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
PS  
XX  
XX Claim 1: Page 289-293; 1409p; English.  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)  
CC recorded on it, or a representative fragment or a sequence at least 958  
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in  
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from  
CC Streptococcus pneumoniae. The present invention also describes an  
CC isolated nucleic acid molecule encoding a homologue of any of the  
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the  
CC nucleic acid molecule is produced by a process comprising: (a) screening  
CC a genomic DNA library using as a probe a target sequence defined by any  
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the  
CC library which contain sequences that hybridise to the target sequence and  
CC isolating the nucleic acid molecules from the members; or (b) isolating  
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid  
CC molecules whose nucleotide sequence is homologous to amplification  
CC primers derived from the fragment of the S. pneumoniae genome to prime  
CC the amplification and isolating the amplified sequences. The computer  
CC readable medium can be used in a computer-based system for identifying  
CC fragments of the S. pneumoniae genome of commercial importance, or  
CC expression modulating fragments of the S. pneumoniae genome. Products  
CC from the present invention can be used in diagnosis kits and assays, and  
CC pharmaceutical compositions and vaccines for S. pneumoniae.  
SQ Sequence 7147 BP; 2088 A; 1641 C; 1265 G; 2153 T; 0 other;  
  
Query Match 20.8%; Score 25.6; DB 19; Length 7147;  
Best Local Similarity 70.8%; Pred. No. 29;  
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
  
CY 28 caaccctatgacctaatgagcgaatgcgagcgctgcgaactgactgta 75  
DB 219 CAACATATGAGGAACATGATGCGCAAGTGGCTTCTACGCTGACTGTA 172  
  
RESULT 17  
AAAT46159  
ID AAAT46159 standard; DNA; 19932 BP.  
XX  
XX AAAT46159;  
XX  
XX 30-JUN-1997 (first entry)  
XX  
XX CagI locus.  
XX  
XX CagI; CagA: virulence factor; exporter molecule; homology; pti gene;  
KW Bordetella pertussis; YJR B4; Agrobacterium tumefaciens; invasion factor;  
KW Salmonella: type I strain; virulence; diagnosis; H. pylori; infection;  
KW vaccine; treatment; duodenal; gastric ulcer; active gastritis;  
KW adenocarcinoma; ss.  
XX  
XX Helicobacter pylori.  
XX  
XX  
XX Key location/Qualifiers  
FH CDS 389..1579  
FT  
FT /\*tag= a  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT CDS 1794..2168  
FT

FT /\*tag= b  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT 3343..4491  
FT  
FT /\*tag= c  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT 4488..5426  
FT  
FT /\*tag= d  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (1178..2173)  
FT  
FT /\*tag= e  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (2591..3001)  
FT  
FT /\*tag= f  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (3020..3259)  
FT  
FT /\*tag= g  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (5508..6233)  
FT  
FT /\*tag= h  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (6218..6613)  
FT  
FT /\*tag= h  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (6562..7074)  
FT  
FT /\*tag= i  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (7007..7375)  
FT  
FT /\*tag= j  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (7371..7802)  
FT  
FT /\*tag= k  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (7975..8454)  
FT  
FT /\*tag= l  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (8496..8918)  
FT  
FT /\*tag= m  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (8977..9762)  
FT  
FT /\*tag= n  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (9784..10575)  
FT  
FT /\*tag= o  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT 10557..12758  
FT  
FT /\*tag= p  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (12743..13420)  
FT  
FT /\*tag= q  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (13374..13742)  
FT  
FT /\*tag= r  
FT /note= "putative open reading frame; no start codon  
FT given"  
FT complement (13866..14129)  
FT  
FT /\*tag= s  
FT



```

FT      /note= "putative open reading frame; no start codon
FT      given"
FT      CDS      complement (19296..19832)
FT      /*tag= t
FT      /note= "putative open reading frame; no start codon
FT      given"
XX      WO9633274-A1.
XX      24-OCT-1996.
XX      18-APR-1996: 96MO-IB00343.
XX      07-JUN-1995: 95US-0477451.
XX      20-APR-1995: 95US-0425194.
XX      (BIOC-) BIOCINE SPA.
XX      Covacci A;
XX      WPI: 1996-485780/48.
XX      P-PSDB: AAM06930-50.
XX      Helicobacter pylori CagI polynucleotide and related proteins - used
XX      in diagnosis and in vaccines for the treatment of H. pylori
XX      infection associated disease
XX      Claim 2: Fig 3A-R: 303pp; English.
XX      The present sequence is the complete nucleotide sequence of the CagI
XX      locus, including the true 5'-terminus of CagI. The sequence was
XX      constructed using overlapping clones (one of which also overlaps with
XX      the Caga region). The CagI region contains clusters of putative open
XX      reading frames (ORFs) with different polarities. The putative ORFs for
XX      this region are shown above. It is hypothesised that some of these ORFs
XX      may encode exporter molecules with homology to the pil genes of
XX      Bordetella pertussis and VIR B4 genes of Agrobacterium tumefaciens and
XX      for proteins with motifs shared by the purported invasion factors of
XX      Salmonella genus. The absence of the Caga gene in the type I strains is
XX      associated with the absence of CagI sequences (which may encode virulence
XX      factors restricted to type I strains). The CagI nucleotide sequence, its
XX      fragments and encoded proteins are used in the diagnosis of H. pylori
XX      (esp. H. pylori type I strain) infection in an individual and in vaccines
XX      (claimed) for the treatment of H. pylori infection associated with e.g.
XX      duodenal and gastric ulcers, severe forms of active gastritis (esp. type
XX      gastritis) and gastric adenocarcinoma.
XX      Sequence 19932 BP; 6677 A; 3635 C; 3483 G; 6137 T; 0 other:
SQ
Query Match      20.8%; Score 25.6; DB 17; Length 19932;
Best Local Similarity 51.8%; Pred. No. 42;
Matches 58; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
OY      11 aaatgtgagatgccttaacaccgagatgactaatggcgcaatgagagcgctgcactac 70
DB      8085 aaatgtgagatgccttaacaccgagatgactaatggcgcaatgagagcgctgcactac 70
OY      71 tggtaagttcacatgtaagtgatctactatccttaattatccacacac 122
DB      8145 cgttaaatccctcgttaattgtctaatattgaagatattgtcccccacagc 8196
RESULT 18
AAZ88746
ID      AAZ88746 standard; DNA: 591 BP.
XX
AC      AAZ88746;
XX
DT      18-MAY-2000 (first entry)
XX
DE      A. dichotoma kabutოსin A DNA.
XX

```

```

KW      Kabutოსin A; kabutოსin B; antibacterial; food preservative;
KW      medical; ds.
XX
OS      Allomyrina dichotoma.
XX
FH      Key      Location/Qualifiers
FT      CDS      82..516
FT      /*tag= a
FT      /product= "Kabutოსin A"
XX
XX      JP2000026499-A.
XX      25-JAN-2000.
XX      13-JUL-1998: 98JP-0197190.
XX      13-JUL-1998: 98JP-0197190.
XX      (NORO ) NORINSUISANSO SANSU KONCHU.
XX      (NODA ) ZH NODA SANGYO KAGAKU KENKYUSHO.
XX      WPI: 2000-176990/16.
XX      P-PSDB: AAY51547.
XX      A new peptide, an antibacterial agent, a new peptide gene, a new
XX      recombinant DNA and preparation of a new peptide
XX      Example 2: Page 9-10; 17pp; Japanese.
XX
XX      This invention describes two novel peptides which have antibacterial
XX      activity, designated kabutოსin A and B and which have been isolated from
XX      Allomyrina dichotoma. The peptides are used as food preservatives
XX      and medical antibacterial agents. This sequence encodes the
XX      A. dichotoma kabutოსin A protein described in the method of the
XX      invention.
XX      Sequence 591 BP; 183 A; 123 C; 119 G; 166 T; 0 other:
SQ
Query Match      20.7%; Score 25.4; DB 21; Length 591;
Best Local Similarity 58.7%; Pred. No. 15;
Matches 44; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY      23 ccttaaacctcgtatgacctaatggcgcaatgagagcgctgcactagtgtaagttca 82
DB      299 ccttaacagcagtgatgcactaatgtcccgatcagccggtcacagttaacctaccagcatca 358
OY      83 catgtaagtgatcac 97
DB      359 cctctaactgtgaac 373
RESULT 19
AAH17997/C
ID      AAH17997 standard; CDNA: 2333 BP.
XX
AC      AAH17997;
XX
DT      26-JUN-2001 (first entry)
XX
DE      Human cDNA sequence SEQ ID NO:17797.
XX
XX      Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX      Homo sapiens.
XX      EP1074617-A2.
XX      07-FEB-2001.
XX      28-JUL-2000; 2000EP-0116126.
XX      29-JUL-1999; 99JP-0248036.
XX

```



PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
DR WPI: 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 17797; 2537bp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13639 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2333 BP; 699 A; 427 C; 416 G; 791 T; 0 other;  
XX  
Query Match 20.7%; Score 25.4; DB 22; Length 2333;  
Best Local Similarity 56.6%; Pred. No. 24;  
Matches 47; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
XX  
OY 27 acaacccatgatgacctaattggcgcaatgagggcgacgactgtaagttcaatg 86  
DB 1547 ACTGCACAGACCCCTGTACTGCGACCAAGTTGATGGTCACTGGAAAAAATAATGCCAATA 1488  
XX  
OY 87 taaggtgatactatctcaatt 109  
DB 1487 TAAAGTCGAAGTCATGATTCAAAT 1465  
XX  
RESULT 20  
AAA92497  
ID AAA92497 standard; DNA; 3300 BP.  
XX  
AC AAA92497;  
XX  
DT 17-JAN-2001 (first entry)  
XX  
DE Haemophilus influenzae adhesin (Hia) gene from NTHI strain K9.  
XX  
KW Hia: adhesin; Haemophilus influenzae adhesin; NTHi: infection; vaccine;  
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;  
KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;  
KW diagnosis; immunogenic; antigen; ds.  
XX

OS Haemophilus influenzae.  
XX  
PN WO200055191-A2.  
XX  
PD 21-SEP-2000.  
XX  
XX 16-MAR-2000; 2000WO-CA00289.  
PF  
XX  
PR 16-MAR-1999; 99US-0268347.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
PI Loosmore SM, Yang Y, Klein MH;  
PI WPI: 2000-618897/59.  
DR P-PSDB: AAB23858.  
XX  
XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
PT use as antigens and vaccines and for treating Hemophilus influenzae  
PT infection -  
XX  
PS Claim 1; Fig 22; 275bp; English.  
XX  
XX The present sequence represents a Haemophilus influenzae adhesin (Hia)  
CC gene from the non-typeable Haemophilus influenzae (NTHi) strain K9.  
CC Hia genes and proteins have antiinflammatory, auditory and antibacterial  
CC activities, and can be used in the production of a vaccine. An  
CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
CC protection against disease caused by Haemophilus strains in a  
CC susceptible host, preferably a human. An Hia protein is useful as an  
CC antigen in immunogenic preparations including vaccines, as a carrier  
CC for other immunogens, and in the generation of diagnostic reagents. Hia  
CC is useful for treating diseases caused by the infection of Haemophilus  
CC influenzae such as meningitis, epiglottitis, septicemia and otitis  
CC media. Recombinant production of Hia favours high recovery of the  
CC protein compared to the low recovery of native protein from Haemophilus  
CC influenzae species. A truncated protein has a significantly higher  
CC amount of recovery than a full-length protein.  
XX  
SQ Sequence 3300 BP; 1099 A; 606 C; 819 G; 776 T; 0 other;  
XX  
Query Match 20.7%; Score 25.4; DB 21; Length 3300;  
Best Local Similarity 61.2%; Pred. No. 27;  
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
XX  
OY 57 agggctgcaactgtaagttcacatgttaagtgatactatctcaattatcca 116  
DB 154 agtgcctctgttaccgagtagtgatgttaaggcgatacgatactaatcaatgca 213  
XX  
OY 117 cacaca 123  
DB 214 gccacata 220  
XX  
RESULT 21  
AAH83942/C  
ID AAH83942 standard; DNA; 649 BP.  
XX  
AC AAH83942;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Callithrix jacchus olfactory receptor encoding gene 10.  
XX  
KW Olfactory receptor; primate; mouse; human; food processing industry;  
KW aromas; perfumery; toxic substance; ds.  
XX  
KW Callithrix jacchus.  
XX  
OS  
XX WO200146262-A2.  
XX



```

PD 28-JUN-2001.
XX
XX 22-DEC-2000; 2000WO-IB02017.
XX
XX 22-DEC-1999; 99US-0171746.
PR 21-DEC-2000; 2000US-0747155.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Rouquier S, Giorgi D;
XX
XX WPI: 2001-381911/40.
DR P-PSDB: AAG98470.
XX
XX Nucleic acids encoding primate and murine olfactory receptors, useful
PT for analysis odours e.g. in food processing and perfumery -
XX
XX Claim 1; Page 150-151; 482pp; English.
XX
XX The invention relates to olfactory receptors (AAG98432-AAG98609) and the
CC genes encoding them (AAH83879-AAH84131) including pseudogenes of 10
CC primate species, mouse and human. The nucleic acids and receptors may be
CC used in the food processing industry (e.g. for the detection of aromas,
CC quality control and sample analysis), in perfumery (e.g. for the analysis
CC or comparison of perfumes) and in the environment (e.g. for the detection
CC of toxic substances and/or trapping of odours).
XX
XX Sequence 649 BP; 126 A; 192 C; 139 G; 192 T; 0 other;
SQ
Query Match 20.5%; Score 25.2; DB 22; Length 649;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 1 caggatgtgaatgtgagatgccttaacacctgaatgagccaatgagcgaggg 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 271 CAGGGAATGTTAGTGGGAGATGCTGCACAGACCAATACATGAGTAGGCAATGCCGGGG 212
QY 61 ctgacactgac 70
    ||| ||| |||
DB 211 GTTCATAGCC 202
DE
XX
XX RESULT 22
XX AAI60262
XX ID AAI60262 standard; cDNA; 3684 BP.
XX
XX AAI60262;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 4251.
XX
XX Human; noctropic; immunosuppressant; cytosstatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.

```

```

PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR P-PSDB: AAM41106.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 4251; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,
CC immunosuppressant and cytosstatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 3684 BP; 970 A; 818 C; 968 G; 928 T; 0 other;
SQ
Query Match 20.5%; Score 25.2; DB 22; Length 3684;
Best Local Similarity 52.9%; Pred. No. 33;
Matches 54; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
QY 21 tgacctaacacctgatacctaataatgagcgagggctgacactgactgtaagttt 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3123 tgcaatgcaaccccccacgtccctctgtccccaagctaccacttaagttcaacttttt 3182
QY 81 cacatgtaagtgatcactatcttcaatttcaatttccacacac 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3183 actagtcaaatagatcttcaactcaataatattccacacac 3224
DE
XX
XX RESULT 23
XX AAT33537
XX ID AAT33537 standard; DNA; 12412 BP.
XX
XX AAT33537;
XX
XX 15-FEB-1998 (first entry)
XX
XX BCG deletion region 3 and flanking sequences.
XX
XX BCG delta 3; virulence; avirulence; attenuation; gene deletion;
KW mycobacteria; vaccine; infection; marker; ss.
XX
XX Mycobacterium bovis strain BCG.
XX
XX Key Location/Qualifiers
XX FT misc_feature 1406..10673
XX FT /tag= a
XX FT /note= "BCG delta 1 deletion region"
XX
XX WO9625519-A1.

```



PD 22-AUG-1996.  
XX  
XX 15-FEB-1996; 96WO-US01938.  
XX  
PR 17-FEB-1995; 95US-0390878.  
XX  
PA (PATH-) PATHOGENESIS CORP.  
XX  
PI Mahairas GG, Stover CK;  
XX  
DR WPI; 1996-393419/39.  
XX  
XX Detecting markers for avirulence in *Mycobacterium* - used in  
PT production of vaccines against bacterial infection, and to detect  
PT bacterial infection  
XX  
XX Example 1; Fig 3; 66pp; English.  
XX  
XX This DNA sequence comprises *Mycobacterium bovis* BCG deletion  
CC sequence BCGdelta3. A specific genetic deletion of this region  
CC results in an avirulence phenotype of the mycobacterium. 2 other  
CC deletion regions (see AAI33535 and AAI33536) have also been detected.  
CC Identification involved screening a BCG cosmid library with a  
CC radiolabeled probe obtained following DNA subtraction between  
CC virulent *Mycobacterium tuberculosis* H37Rv and avirulent BCG.  
CC The deletions provide useful markers for the identification of an  
CC avirulent, or a virulent, mycobacterial phenotype. Determination  
CC of avirulence requires the detection of the presence or absence of  
CC the deletion; the deletions are detected either by detecting the  
CC presence or absence of deletion junctions (see AAI33538-46), or by  
CC detecting the presence or absence of the sequences contained within  
CC the deletion. Deletion polypeptides are used as components of  
CC immunological assays and in vaccines.  
XX  
SQ Sequence 12412 BP; 2146 A; 4273 C; 3876 G; 1917 T; 200 other;  
  
Query Match 20.5%; Score 25.2; DB 17; Length 12412;  
Best Local Similarity 66.7%; Pred. No. 50;  
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
  
QY 17 gagatgccttacacacctgatgacctaatgacgaatgagggctgacctgac 70  
DB 5990 gacatgcttcaacgcgcgatgctgtaigcgacctgcaaacgctgtgtgac 6043  
  
RESULT 24  
AAH32402/C  
ID AAH32402 standard; DNA; 933 BP.  
XX  
AC AAH32402;  
XX  
DT 30-JUL-2001 (first entry)  
XX  
XX Human olfactory receptor polynucleotide, SEQ ID NO: 975.  
DE  
XX Human olfactory receptor; OR: primary scent determination;  
KM secondary scent determination; polypeptide library; odour receptor;  
KW scent profile; scent fingerprint; scent representation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200127158-A2.  
XX  
PD 19-APR-2001.  
XX  
XX 06-OCT-2000; 2000WO-US27582.  
PF  
XX 08-OCT-1999; 99US-0158615.  
PR  
XX 24-FEB-2000; 2000US-0184809.  
XX  
PA (DIGI-) DIGISCENTS.  
XX  
PA (YEDA ) YEDA RES & DEV CO LTD.

XX  
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;  
XX  
XX WPI; 2001-290713/30.  
DR  
XX  
XX New polynucleotides which encode polypeptides involved in olfactory  
PT sensation for identifying olfactory agonists and antagonists -  
PT  
XX  
PS Claim 8; Page 569; 1857pp; English.  
XX  
XX The present sequence is one of a number of isolated polynucleotides  
CC which encode polypeptides involved in olfactory sensation. The  
CC polynucleotides can be used in screening for olfactory agonists and  
CC antagonists. The methods allow for the determination of primary  
CC scents and the identification of the odour receptors used to detect  
CC these primary scents. The methods also enable determination of  
CC secondary scents and the identification of combinations of odour  
CC receptors that are involved in detecting such secondary scents.  
CC This enables the construction of a scent representation (also called  
CC a scent fingerprint or scent profile), which may be used to re-create  
CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors,  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.  
XX  
SQ Sequence 933 BP; 206 A; 267 C; 187 G; 273 T; 0 other;  
  
Query Match 20.3%; Score 25; DB 22; Length 933;  
Best Local Similarity 61.5%; Pred. No. 24;  
Matches 40; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
  
QY 1 caggtatgtaaatgtagatgcttaccacacctgacctgaatgacgaatgaggg 60  
DB 474 CAGGGAATGTAGGTGGAGATGCTGCAGAACCAATGATGATGATGCCATGCG 415  
QY 61 ctgca 65  
DB 414 TTTC A 410  
  
RESULT 25  
AAAX04441/C  
ID AAAX04441 standard; cDNA; 2284 BP.  
XX  
AC AAAX04441;  
XX  
DT 27-APR-1999 (first entry)  
XX  
XX S. frugiperda immunophilin FKBP46 cDNA.  
DE  
XX Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;  
KW transplant; tissue graft; ds.  
XX  
XX Spodoptera frugiperda.  
OS  
XX  
FH Key location/Qualifiers  
FT CDS 160..1398  
FT /\*tag= a  
FT /\*product= "immunophilin FKBP46"  
FT 1764..1769  
FT /\*tag= b  
FT polya\_signal 1850..1855  
FT /\*tag= c  
FT polya\_signal 1868..1873  
FT /\*tag= d  
FT polya\_signal 2052..2057  
FT /\*tag= e  
FT polya\_signal 2237..2242  
FT /\*tag= f  
XX  
XX US5861498-A.  
XX  
XX







PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 3; SEQ ID 9963; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 528 BP; 123 A; 126 C; 127 G; 132 T; 20 other;  
  
Query Match 20.2%; Score 24.8; DB 22; Length 528;  
Best local Similarity 60.3%; Pred. No. 23;  
Matches 41; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
  
QY 31 ccctgatgacctatgacgaatgcagggctgactgactgtaagtacatgtaag 90  
DB 95 CCTCGAAGAGCTGTGCTCAGCAGCGGCTTCAAGGGATGGGAACCTTTATTTCAAA 36  
QY 91 gtgtact 98  
DB 35 TAAACTACT 28  
  
RESULT 28  
AAC39501/c  
ID AAC39501 standard; DNA; 1448 BP.  
XX  
AC AAC39501;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 24864.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
KM metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134270.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.











Query Match	20.2%	Score 24.8;	DB 16;	Length 2064;
Best Local Similarity	57.9%	Pred. No. 37;		
Matches 44; Conservative	0;	Mismatches 32;	Indels 0;	Gaps 0

RESULT	32
AAQ80533	
ID	AAQ80533 standard; cDNA to mRNA; 2064 BP.

XX Paralicthys olivaceus.  
OS  
KW transglutaminase; fish; recombinant production; Pagrus major;  
KM Theragra chalcogramma; Paralicthys olivaceus; Oncochinchus keta; ds

PF	13-JUL-1993;	93JP-0172998.
XX		
PR	13-JUL-1993;	93JP-0172998.

PT A fish-originated trans:glutaminase gene - can be produced in large  
PT amounts at low economic cost.

CC The DNA encodes a protein, isolated from *Paralichthys olivaceus*, which  
CC has transglutaminase (TG) activity. The TG's of the invention can be

Query Match	20.28;	Score 24.48;	DB 16;	Length 2064;
Best Local Similarity	57.98;	Pred. No. 37;		
Matches 44; Conservative	0;	Mismatches 32;	Indels 0;	Gaps 0

QY	68	gactggtgaagtttcac	83
Db	693	gactgtcagtggcac	708

AC	AAQ46026;
XX	
DT	21-FEB-1994 (first entry)
VV	

KW Transglutaminase; TGase; fish; Pagrus major; Alaska pollack;  
KW Theragra chalcogramma; Paralichthys olivaceus; paste; kamaboko,  
KW gelatinous food; cosmetic; ss.

key	Location/Qualifiers
PH	26..2092
FT	/*tag= a
FT	

PF	14-JAN-1993;	93EP-0100488
XX		
PR	14-JAN-1992;	92JP-0005166
PD	07-JUL-1993;	93JD-0100903

PR 08-DEC-1992; 92Jp-0328010.

PI Matsui H., Motoki M., Nagase K., Nakanishi K., Yasueda H.  
xx  
xx

PT New gene encoding trans:glutaminase from fish - for the mass  
PT prodn. of trans:glutaminase for use in gelatinous food, cosmetics  
PT and chemicals

CC The difference between the sequences of AAQ46025 and AAQ54878 residues  
CC in the 1882 th base. The change does not alter the translated

CC Transglutaminase is used in the prodn. of gelatinous food prods.,  
CC cosmetics and chemicals. The DNA fragment enables the mass prodn  
CC of transglutaminase which is cheap and efficient.

**SQ** Sequence 2148 BP; 578 A; 496 C; 610 G; 464 T; 0 other;



	Query Match	Similarity	20.2%	Score 24.8;	DB 14;	Length 2148;	
	Best Local Similarity	57.9%;	Pred. No.	38;	Mismatches	32;	Indels 0; Gaps 0;
Oy	Matches	44;	Conservative	0;	Mismatches	32;	Indels 0; Gaps 0;
Db	661	tgttaattgtagatgccttacacacctgatgaactaatgycgaalgcagggcgtcac	67            				
Oy	68	gactcgtaaagtccac	83 				
Db	721	gactgttcagttgccac	736				
	RESULT_34						
ID	AAQ54878	standard; CDNA to mRNA; 2148 BP.					
AC	AAQ54878;						
XX	21-FEB-1994	(first entry)					
DE	Transglutaminase (fish liver) (G1882).						
KW	Transglutaminase; rGase; fish; Pagrus major; Alaska pollack; Theragra chalcogramma; Paralichthys olivaceus; paste; kamaboko; gelatinous food; cosmetic; ss.  Paralichthys olivaceus.						
FH	Key Location/Qualifiers						
FT	CDS / *tag= a						
NN	EP555649-A.						
PD	18-AUG-1993.						
PF	14-JAN-1993; 93EP-0100488.						
PR	14-JAN-1992; 92JP-0005166. 27-JUL-1992; 92JP-019803. 08-DEC-1992; 92JP-0328010.						
PA	(AJIN ) AJINOMOTO KK.						
PI	Matsui H, Motoki M, Nagase K, Nakanishi K, Yasueda H;						
DR	WPJ: 1993-259983/33.						
PS	P-PSTB; AAR39519.						
PT	New gene encoding trans:glutaminase from fish - for the mass prodn. of trans:glutaminase for use in gelatinous food, cosmetics and chemicals						
CS	Claim 1; Page 45-47; 85pp; English.						
CC	The difference between the sequences of AAQ546025 and AAQ54878 resides in the 1882 th base. The change does not alter the translated prod.						
CC	Transglutaminase is used in the prodn. of gelatinous food prods., cosmetics and chemicals. The DNA fragment enables the mass prodn. of transglutiminase which is cheap and efficient.						
SQ	Sequence 2148 BP: 578 A; 496 C; 611 G; 463 T; 0 other:						
	Query Match	20.2%; Score 24.8;	DB 14;	Length 2148;			
	Best Local Similarity	57.9%; Pred. No.	38;	Mismatches	32;	Indels 0;	Gaps 0;
Oy	8	tgttaattgtagatgccttacacacctgatgaactaatgycgaalgcagggcgtcac	67            				
Db	661	tgtctaattgcagaagatacatcaaccgcatgttgtaacctaacggtgacaagggttgttgc	720         				

Oy	68	gacttgtaattccac	83
Db	721	gactgtgcagtcggcac	736
RESULT	35		
AAQ80536			
ID	AAQ80536	standard; cDNA to mRNA; 2148 BP.	
XX			
AC	AAQ80536;		
DT	31-MAY-1996	(first entry)	
XX			
DE	Paralichthys olivaceus transglutaminase gene.		
XX			
RW	transglutaminase; fish; recombinant production; Pagrus major;		
KW	Theragra chalcogramma; Paralichthys olivaceus; Oncorhynchus keta; ds.		
XX			
OS	Paralichthys olivaceus.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	26..2092	
FT		/tag= a	
FT		/product= transglutaminase	
XX			
PN	JF07023787-A.		
XX			
PD	27-JAN-1995.		
XX			
XX	13-JUL-1993;	93JP-0172998.	
PF			
XX			
PR	13-JUL-1993;	93JP-0172998.	
XX			
PA	(AJIN ) AJINOMOTO KK.		
XX			
XX	WPI; 1995-100948/14.		
DR	P-PSDB; AAR65331.		
XX			
PT	A fish-originated trans:glutaminase gene - can be produced in large		
XX	amounts at low economic cost.		
XX			
XX			
P4	Disclosure; Page 54-56; 94pp; Japanese.		
XX			
CC	The DNA encodes a protein, isolated from Paralichthys olivaceus, which		
CC	has transglutaminase (TG) activity. The TG's of the invention can be		
CC	isolated from Pagrus major, Theragra chalcogramma, Paralichthys		
CC	olivaceus or Oncorhynchus keta. Fish-originated TG's are economical		
CC	to produce in large quantities.		
XX			
XX			
SQ	Sequence 2148 BP; 578 A; 496 C; 611 G; 463 T; 0 other:		
Query Match	20.2%;	Score 24.8;	DB 16; Length 2148;
Best Local Similarity	57.9%;	Pred. No. 38;	
Matches	44;	Conservative 0;	Mismatches 32; Indels 0; Gaps 0;
OY	8	tgtaaatgtggaagccttacaccctgatgacctaatggcgcaatggagggtgcact	67
Dt	661	tgtctatgtcgcaggaatcaccctgcagtgttgtaacctaacggtgacaagggtgtgct	720
OY	68	gactgtaagtctcac	83
Db	721	gactgttcagtcggcac	736
RESULT	36		
AAQ80534			
ID	AAQ80534	standard; cDNA to mRNA; 2148 BP.	
XX			
AC	AAQ80534;		
DT	31-MAY-1996	(first entry)	



```

XX DE Parallelthys olivaceus transglutaminase gene.
XX DE transglutaminase; fish; recombinant production; Pagrus major;
XX KW Theragra chalcogramma; Parallelthys olivaceus; Oncorhynchus keta; ds.
XX OS
XX DE Parallelthys olivaceus.
XX DE
XX FH Key Location/Qualifiers
XX FT CDS 26..2092
XX FT /tag= a
XX FT /product= transglutaminase
XX
XX JF07023787-A.
XX
XX PD 27-JAN-1995.
XX
XX PF 13-JUL-1993; 93JP-0172998.
XX
XX PR 13-JUL-1993; 93JP-0172998.
XX
XX (AJIN ) AJINOMOTO KK.
XX
XX WPI: 1995-100948/14.
XX
XX P-PSDB: AAR65331.
XX
XX A fish-originated trans:glutaminase gene - can be produced in large
XX PT amounts at low economic cost.
XX
XX Disclosure: Page 50-53; 94pp; Japanese.
XX
XX The DNA encodes a protein, isolated from Parallelthys olivaceus, which
XX CC has transglutaminase (TG) activity. The TG's of the invention can be
XX CC isolated from Pagrus major, Theragra chalcogramma, Parallelthys
XX CC olivaceus or Oncorhynchus keta. Fish-originated TG's are economical
XX CC to produce in large quantities.
XX
XX SO Sequence 2148 BP; 578 A; 496 C; 610 G; 464 T; 0 other;

Query Match 20.2%; Score 24.8; DB 16; Length 2148;
Best Local Similarity 57.9%; Pred. No. 38;
Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 8 tgaatgtgagatgccttaacaacctgatagcctaagcgcaatgcgaggtgcac 67
Db 661 tgcatagtcagcagatgatactcgtgtaacttaacgagcgaggtgtgct 720
QY 68 gactgtgaatttcac 83
Db 721 gactgtgactgtggtcac 736

RESULT 37
AAC59911
ID AAC59911 standard; cDNA; 452 BP.
XX
XX AAC59911:
XX
XX 30-JAN-2001 (first entry)
XX
XX Human secreted protein cDNA sequence #5.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX KW vunerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein; ss.
XX
XX Homo sapiens.
XX
XX WO200055198-A1.
XX

```

```

PD 21-SEP-2000.
XX
XX PF 09-MAR-2000; 2000WO-US06012.
XX
XX PR 12-MAR-1999; 99US-0124093.
XX
XX PR 23-NOV-1999; 99US-0166989.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI: 2000-587520/55.
XX
XX P-PSDB: AAB28706.
XX
XX Human secreted proteins and the nucleic acids that encode them, useful
XX PT in gene therapy protocols and recombinant nucleic acid based procedures
XX
XX Claim 1; Page 315; 391pp; English.
XX
XX The invention relate to the isolation of genes AAC59907-C59956 encoding
XX CC 50 human secreted proteins AAB28702-B28751. The genes can be used to
XX CC generate fusion proteins by linking to the gene for the human
XX CC immunoglobulin G Fc portion for increasing the stability of
XX CC the fusion protein as compared to the human protein only. The genes and
XX CC proteins are useful for preventing, ameliorating or treating medical
XX CC conditions, e.g. by protein or gene therapy. The genes are isolated
XX CC from a range of human tissues disclosed in the specification. The
XX CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
XX CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX CC and parasitic infections.
XX
XX SO Sequence 452 BP; 116 A; 104 C; 101 G; 128 T; 3 other;

Query Match 20.0%; Score 24.6; DB 21; Length 452;
Best Local Similarity 65.5%; Pred. No. 26;
Matches 36; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 67 tgactgtgaatttcacatgataagtgatactatcttaatttccacaaa 121
Db 209 tcactgttcagtgctactgtttgtaacatacaatcaacgtagtctaccaa 263

RESULT 38
AAF26556
ID AAF26556 standard; DNA; 625 BP.
XX
XX AAF26556:
XX
XX 27-MAR-2001 (first entry)
XX
XX DNA encoding human secreted protein #10.
XX
XX Secreted protein; gene therapy; vaccine; cancer; leukemia;
XX KW autoimmune disease; allergy; inflammation; graft rejection;
XX KW hyperproliferation; cardiovascular; infection; ss.
XX
XX Homo sapiens.
XX
XX WO200076531-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US15137.
XX

```







CC transgenic viral resistant shrimp.

SO Sequence 2946 BP; 979 A; 566 C; 665 G; 736 T; 0 other;

Query Match 20.0%; Score 24.6; DB 22; Length 2946;  
Best Local Similarity 70.2%; Pred. No. 50;  
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 69 actggaattcacatgaagtgatactatctcaattatcc 115  
Db 1549 ACTCGATTTCACATTCATTCAACTTACCTCACTCTCTCC 1503

RESULT 41  
AAH81775  
ID AAH81775 standard; DNA: 3787 BP.

XX AAH81775;

XX 21-SEP-2001 (first entry)

XX Human differential transcription-associated cDNA SEQ ID 284.

XX Differential transcription; human; rat; tumour cell; cytostatic;

XX Ras modulator; Class II tumour suppressor gene; gene therapy; ss.

XX Homo sapiens.

XX W0200157058-A2.

XX 09-AUG-2001.

XX 31-JAN-2001; 2001WO-EP01003.

XX 31-JAN-2000; 2000DE-1004102.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;

XX Grips M, Hellriegel M, Schmitz A, Sers C;

XX WPI; 2001-483415/52.

XX Nucleic acids differentially expressed between tumor and normal cells,  
XX useful for diagnosis or therapy of tumors and for screening active  
XX agents

PS Disclosure; Page 438-439; 579pp; German.

XX This invention describes a nucleic acid (I) with differential expression  
XX between tumour and normal cells and which has cytostatic activity. (I)  
XX work as modulators of Ras activity by inducing expression of tumour  
XX suppressor genes. (I), and polypeptides encoded by them, are useful as  
XX targets for diagnosis or therapy and in screening to determine the  
XX effects of an active compound (potential pharmaceutical) on a cell line,  
XX particularly for diagnosis and treatment of tumors, especially by  
XX modulating expression of (I) (by gene therapy, antisense RNA or ribozyme  
XX methods) or by modulating the amount and/or location of (I)-encoded  
XX polypeptides (by administration of the polypeptide or its activator,  
XX antibody (optionally as a conjugate) or inhibitor). The method allows  
XX identification of many Class II tumour suppressor genes (i.e. genes that  
XX are not primary targets for tumour-initiating mutations).  
XX AAH81492-AAH82376 represent the human and rat derived nucleic acid  
XX fragments described in the method of the invention.

SO Sequence 3787 BP; 898 A; 1210 C; 1073 G; 606 T; 0 other;

Query Match 20.0%; Score 24.6; DB 22; Length 3787;  
Best Local Similarity 59.2%; Pred. No. 54;  
Matches 42; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 gtatgttaatgtgagatgctctacacccgtatgaccctaatggcgcaatgagagctg 63  
Db 372 gtactgcgtctgccgcgtccctacagcgttaccgccttatgcagtcgacgtgctg 431

QY 64 cactgactggt 74  
Db 432 caagagactggt 442

RESULT 42  
AAK57351  
ID AAK57351 standard; DNA: 68940 BP.

XX AAK57351;

XX 16-AUG-1999 (first entry)

XX Human chromosome 6q27 tumour suppressor gene.

XX Tumour suppressor gene; chromosome 6q27; cancer; therapy; human; D6S297;

XX breast cancer; ovarian cancer; lymphoid cancer; marker; D6S193; ss.

XX Homo sapiens.

XX Key location/Qualifiers

XX exon 414..549

XX polyA\_site /tag= a complement (719..724)

XX repeat\_region /tag= b 1126..1220

XX variation /tag= c /note= "MIR repeat family"

XX variation /tag= d replace(1128,T)

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /tag= e replace(1909,C)

XX variation /note= "putative polymorphism, A in 52b3, C in

XX variation /tag= f replace(1920,T)

XX exon /note= "putative polymorphism, C in 52b3, T in

XX exon /tag= g complement (2310..2482)

XX exon /tag= h 2749..2843

XX variation /tag= i replace(2984,T)

XX exon /note= "putative polymorphism, C in 52b3, T in

XX exon /tag= j complement (3083..3201)

XX exon /note= "trapped exon"

XX promoter /tag= k 3654..3793

XX promoter /tag= l complement (4483..4532)

XX promoter /tag= m complement (4615..4664)

XX promoter /tag= n 4854..4903

XX promoter /tag= o 4878..4927

XX exon /tag= p 4912..5093

XX misc\_feature /tag= q complement (5398..5563)

XX promoter /tag= r /note= "match: D6S193 microsatellite"

XX promoter /tag= s complement (5558..5607)

XX promoter 5767..5816



```

FT      /*tag= s
FT      6072..6134
FT      /*tag= t
FT      complement (6395..6438)
FT      exon
FT      /*tag= u
FT      /note= "trapped exon"
FT      complement (6395..6456)
FT      /*tag= v
FT      complement (6887..6892)
FT      polyA_site
FT      /*tag= w
FT      complement (7237..7434)
FT      exon
FT      /*tag= x
FT      complement (8055..8165)
FT      exon
FT      /*tag= y
FT      complement (8414..8447)
FT      /*tag= z
FT      8530..8695
FT      /*tag= aa
FT      complement (8543..8593)
FT      exon
FT      /*tag= ab
FT      /note= "trapped exon"
FT      9351..9678
FT      exon
FT      /*tag= ac
FT      /note= "trapped exon"
FT      complement (9467..9583)
FT      exon
FT      /*tag= ad
FT      9627..9678
FT      /*tag= ae
FT      complement (9636..9685)
FT      promoter
FT      /*tag= af
FT      complement (10423..10461)
FT      exon
FT      /*tag= ag
FT      complement (10669..10761)
FT      exon
FT      /*tag= ah
FT      complement (10706..10761)
FT      exon
FT      /*tag= ai
FT      11067..11368
FT      repeat_region
FT      /note= "Alu repeat family"
FT      /*tag= aj
FT      11264..11315
FT      exon
FT      /*tag= ak
FT      complement (12270..12275)
FT      polyA_site
FT      /*tag= al
FT      complement (12641..12690)
FT      promoter
FT      /*tag= am
FT      13800..14074
FT      exon
FT      /*tag= an
FT      /note= "trapped exon"
FT      complement (14197..14400)
FT      exon
FT      /*tag= ao
FT      complement (14496..14601)
FT      repeat_region
FT      /*tag= ap
FT      /note= "L1 repeat family"
FT      complement (14607..14860)
FT      exon
FT      /*tag= aq
FT      /note= "trapped exon"
FT      14930..14979
FT      promoter
FT      /*tag= ar
FT      replace(15091..G)
FT      variation
FT      /*tag= as
FT      /note= "putative polymorphism, A in 52b3 and
FT      pdj84915, G in 255h6"
FT      complement (15609..15614)
FT      polyA_site
FT      /*tag= at
FT      16174..16451
FT      repeat_region
FT      /*tag= au
FT      /note= "Alu repeat family"
FT      16509..16559
FT      exon
FT      /*tag= av
FT      complement (16867..16939)
FT      exon
FT      /*tag= aw
FT      /note= "trapped exon"
FT      16867..17013
FT      exon

```

```

FT      /*tag= ax
FT      /note= "trapped exon"
FT      17288..17352
FT      /*tag= ay
FT      complement (17427..17606)
FT      exon
FT      /*tag= az
FT      17647..17696
FT      /*tag= ba
FT      complement (18367..18416)
FT      promoter
FT      /*tag= bb
FT      complement (19301..19334)
FT      exon
FT      /*tag= bc
FT      complement (19541..19683)
FT      repeat_region
FT      /*tag= bd
FT      /note= "MER20 repeat family"
FT      19688..19975
FT      repeat_region
FT      /*tag= be
FT      /note= "Alu repeat family"
FT      20108..20226
FT      exon
FT      /*tag= bf
FT      replace(20729..20767)
FT      variation
FT      /*tag= bg
FT      /note= "putative polymorphism (TAT13) in 52b3"
FT      complement (20770..20878)
FT      repeat_region
FT      /*tag= bh
FT      /note= "Alu repeat family"
FT      21292..21341
FT      promoter
FT      /*tag= bi
FT      21450..21499
FT      promoter
FT      /*tag= bj
FT      complement (22392..22441)
FT      exon
FT      /*tag= bk
FT      22856..22956
FT      exon
FT      /*tag= bl
FT      complement (23196..23233)
FT      exon
FT      /*tag= bm
FT      complement (23288..23293)
FT      polyA_site
FT      /*tag= bn
FT      replace(24182..24186)
FT      variation
FT      /*tag= bo
FT      /note= "putative polymorphism (T7)5 in pdj84915"
FT      complement (24244..24293)
FT      promoter
FT      /*tag= bp
FT      24266..24315
FT      promoter
FT      /*tag= bq
FT      complement(24600..24649)
FT      promoter
FT      /*tag= br
FT      30550..31665
FT      exon
FT      /*tag= bs
FT      30550..30863
FT      exon
FT      /*tag= bt
FT      32483..32488
FT      polyA_site
FT      /*tag= bu
FT      32483..32488
FT      repeat_region
FT      /*tag= bv
FT      33639..33870
FT      repeat_region
FT      /*tag= bw
FT      /note= "Alu repeat family"
FT      complement (33782..33867)
FT      exon
FT      /*tag= bx
FT      complement (34638..34996)
FT      repeat_region
FT      /*tag= by
FT      /note= "TIGER2 repeat family"
FT      complement (34879..34941)
FT      exon
FT      /*tag= bz

```

```

Query Match      20.0%; Score 24.6; DB 20; Length 68940;
Best Local Similarity 52.4%; Pred. No. 1.5e+02;
Matches 54; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
Oy      14  tgtgagatgccttacacacctgagcctaattgacgaatgcgaaggctgacactgag 73
            ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      48374  tgaataatgacttaagaattatgctcaataatcatgcatgcggtgctcccaagaagatg 48433

```







```
FT CDS complement (38892..40163)
FT /tag= au
FT /product= "EvrS"
FT CDS complement (40216..40890)
FT /tag= av
FT /product= "EvrT"
FT CDS complement (40899..40902)
FT /tag= aw
FT complement (40887..41576)
FT /tag= ax
FT /product= "EvrU"
FT CDS complement (41679..42707)
FT /tag= ay
FT /product= "EvrV"
FT CDS complement (42714..42717)
FT /tag= az
FT complement (42810..43799)
FT /tag= ba
FT /product= "EvrW"
FT CDS complement (43807..43811)
FT /tag= bb
FT complement (43799..44866)
FT /tag= bc
FT /product= "EvrX"
FT CDS complement (45014..45760)
FT /tag= bd
FT /product= "EvrY"
FT CDS complement (45767..45770)
FT /tag= be
FT complement (45962..46714)
FT /tag= bf
FT /product= "EvrZ"
FT CDS complement (45952..45956)
FT /tag= bg
FT complement (47156..49234)
FT /tag= bh
FT /product= "EvsA"
FT CDS 51627..52715
FT /tag= bi
FT /product= "EvsB"
FT CDS 51629..51622
FT /tag= bj
FT 52889..53557
FT /tag= bk
FT /product= "EvsC"
FT CDS 53554..54207
FT /tag= bl
FT /product= "EvsD"
FT CDS complement (54362..55117)
FT /tag= bm
FT /product= "EvsB"
FT CDS complement (55125..55128)
FT /tag= bn
FT CDS complement (55135..56094)
FT /tag= bo
FT /product= "EvsC"
FT CDS complement (56100..56103)
FT /tag= bp
FT CDS complement (56184..56813)
FT /tag= bq
FT /product= "EvsC2"
FT CDS 56961..58709
```

```
Query Match 20.0%; Score 24.6; DB 22; Length 109519;
Best Local Similarity 65.5%; Pred. No. 1.7e+02;
Matches 36; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
OY 18 agatgcttacacacgtatgacctaattggcgaatgcgaagctgcactgactg 72
DB 76022 AGATGGGTAACGACTGGACGACCCGATGCGCCACGCGAGGTCGAGGTCGG 75968
```

RESULT 44

```
AACT5339/C
ID AACT5339 standard; cDNA; 294 BP.
XX
AC AACT5339;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF894 polynucleotide sequence SEQ ID NO:1787.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipariatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihypertic;
KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN MO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000MO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR MPI; 2000-602362/57.
XX
P-PSDB; AAB41130.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 1399; 5507pp; English.
XX
CC AACT74446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipariatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antihypertic; and antinaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 294 BP; 96 A; 51 C; 68 G; 79 T; 0 other;
```



Query Match	Similarity	19.8%	Score 24.4	DB 21	Length 294
Best Local	Similarity	54.4%	Pred. No. 27		
Matches	49	Conservative	0	Mismatches	41
				Indels	0
				Gaps	0

  

Oy	23	ccttaaacacctgatgacctaatgagcgaatgagagctgactgactggaatttca	82
Db	136	cccttataaaacccggaagtccttgcagagatgattaaagtgttcacatc	77
Oy	83	catgtaaggtgatactatctcaatttt	112
Db	76	cttgtagtgtagccctacatttaattcttt	47

  

RESULT	45
AAA01147/c	
ID	AAA01147 standard; cDNA; 300 BP.
XX	
AC	AAA01147;
XX	
DT	19-MAY-2000 (first entry)
XX	
DE	Human colon cancer cell line polynucleotide sequence SEQ ID NO:1138.
XX	
KW	Human; colon cancer; tumour; diagnosis; gene expression product;
KW	probe; detection; cancerous state; metastasis; identification;
KW	breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW	oestrogen receptor-negative breast cancer; lung cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9558675-A2.
XX	
PD	18-NOV-1999.
XX	
PF	13-MAY-1999; 99WO-US10602.
XX	
PR	14-MAY-1998; 98US-0085426.
PR	15-MAY-1998; 98US-0085537.
PR	15-MAY-1998; 98US-0085696.
PR	21-OCT-1998; 98US-0105234.
PR	27-OCT-1998; 98US-0105877.
XX	
PA	(CHIR) CHIRON CORP.
XX	
PI	(HSE-) HYSEQ INC.
XX	
PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX	
PI	Reinhardt C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX	
PI	Lamson G, Dirmanc R, Cirkvenjakov R, Dickson M, Dirmanc S, Labat I;
XX	
PI	Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX	
DR	WPI: 2000-126369/11.
XX	
XX	Polynucleotide library used to determine cancerous states of mammalian
PT	cells -
XX	
PS	Claim 1; Page 454; 1097pp; English.
XX	

  

AAA00010 to AAA02716 represent polynucleotides isolated from cDNA libraries constructed from human colon cancer cell lines. The present invention also describes a method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, where detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived. The polynucleotide sequences can be used in a method for detecting differentially expressed genes correlated with a cancerous state of a mammalian cell. The polynucleotides can also be used as probes for detecting and mapping related genes. They can be used in diagnosis and prognosis of diseases and disorders (e.g. identification of pre-metastatic or metastatic cancerous states, stages of cancer, or

CC responsiveness of cancer to therapy). This is particularly for breast  
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
CC negative breast cancer, lung cancer, and colon cancer.  
XX  
SQ Sequence 300 BP; 85 A; 72 C; 73 G; 70 T; 0 other;

Query Match	19.8%	Score 24.4;	DB 21;	Length 300;
Best Local Similarity	63.8%	Pred. No. 27;		
Matches 37;	Conservative	0;	Mismatches 21;	Indels 0;
Gaps	0;			

  

Oy	1	caggtatgtgaatgtcgagatgcocttaaacctbatacctaattggcgaatgcgag	58
Db	269	CACTATTGGAACCTTAATCTCAATGCACACAGTATATGGAGATGGCCCAATGGAG	212

Search completed: March 19, 2002, 10:21:41  
Job time: 2850 sec







**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:31:46 ; Search time 39 Seconds

(without alignments)  
714.277 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 caggtatgtaaatgtaga.....tcaattatccacacaca 123

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCITUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25.6	20.8	5599	2	US-08-477-451-9
2	25.6	20.8	5599	2	US-08-477-451-13
3	25.6	20.8	19932	2	US-08-477-451-25
4	25.2	20.5	12412	1	US-08-390-878-18
5	25	20.3	2255	2	US-08-741-134-1
6	24.8	20.2	2064	1	US-08-164-839-30
7	24.8	20.2	2064	1	US-08-164-839-32
8	24.8	20.2	2064	1	US-08-583-799-30
9	24.8	20.2	2064	1	US-08-583-799-32
10	24.8	20.2	2148	1	US-08-164-839-69
11	24.8	20.2	2148	1	US-08-164-839-71
12	24.8	20.2	2148	1	US-08-583-799-69
13	24.8	20.2	2148	1	US-08-583-799-71
14	24.4	19.8	7720	4	US-09-318-448-5
15	24.2	19.7	1839	3	US-08-482-677-1
16	24.2	19.7	1839	3	US-08-482-677-5
17	24.2	19.7	6642	1	US-08-727-034-5
18	24.2	19.7	6843	1	US-08-727-034-6
19	23.6	19.2	2329	1	US-08-253-785-2
20	23.4	19.0	562	2	US-08-975-316-53
21	23.4	19.0	1026	2	US-08-713-000-6
22	23.4	19.0	1026	2	US-08-975-316-6
23	23.4	19.0	1026	4	US-09-211-710-6
24	23	18.7	1959	4	US-09-201-664-1
25	23	18.7	2059	1	US-08-244-686-1
26	23	18.7	2277	1	US-08-369-796-7
27	23	18.7	2277	2	US-08-852-091-7

c 28	23	18.7	2277	2	US-08-820-754-7	Sequence 7, Appl
c 29	23	18.7	2277	3	US-08-956-652-7	Sequence 7, Appl
c 30	23	18.7	2277	3	US-08-956-865-7	Sequence 7, Appl
c 31	23	18.7	2277	3	US-08-948-547-7	Sequence 7, Appl
c 32	23	18.7	2277	5	PCT-US95-17025-7	Sequence 7, Appl
c 33	23	18.7	3464	4	US-09-318-448-30	Sequence 30, Appl
c 34	23	18.7	4233	4	US-09-056-105-27	Sequence 27, Appl
c 35	23	18.7	4440	3	US-07-792-600-1	Sequence 1, Appl
c 36	23	18.7	4440	3	US-09-157-021-1	Sequence 1, Appl
c 37	23	18.7	4430	3	US-09-156-842-1	Sequence 1, Appl
c 38	23	18.7	5433	3	US-09-157-021-35	Sequence 35, Appl
c 39	23	18.7	5433	3	US-09-156-842-35	Sequence 35, Appl
c 40	22.8	18.5	180	3	US-08-483-503A-2	Sequence 2, Appl
c 41	22.8	18.5	579	2	US-08-896-501A-7	Sequence 5, Appl
c 42	22.8	18.5	579	2	US-08-896-501A-5	Sequence 5, Appl
c 43	22.8	18.5	680	4	US-09-328-111-774	Sequence 774, App
c 44	22.8	18.5	911	4	US-08-943-731-81	Sequence 81, Appl
c 45	22.8	18.5	1120	3	US-08-884-324-2	Sequence 2, Appl

## ALIGNMENTS

```
RESULT 1
US-08-477-451-9
; Sequence 9, Application US/08477451
; Patent No. 5928665
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5599 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-477-451-9

Query Match 20.8%; Score 25.6; DB 2; Length 5599;
Best Local Similarity 51.8%; Pred. No. 5.5;
Matches 58; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 11 aaatggaatgccttaacacctgatactgacgagcgaatgagggcgtgacatgac 70
Db 4528 AATAGCTTCACATCTGATTAACAGCACCATTTGGCGCGCTTCGTCGTCCTTGTCG 4587
QY 71 tggtaagttcacatgtgaagatatactatctcaattatccacacaac 122
```



Db 4588 CGTTAAATTCCTGTTAATTGCTAAATATTAGATATTGTTCCACACG 4639

## RESULT 2

US-08-477-451-13/C

; Sequence 13, Application US/08477451  
; Patent No. 5928865  
; GENERAL INFORMATION:  
; APPLICANT: Covacci, Antonello  
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,451  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0335.002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-601-2708  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5599 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-477-451-13

Query Match 20.8%; Score 25.6; DB 2; Length 5599;  
Best Local Similarity 51.8%; Pred. No. 5.5;  
Matches 58; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 11 aaatgtagatgaccttacaacctgatacctaataatgagcgaatgagggctgcaactgac 70  
Db 1072 AAATGCTTGCAATCTGACTATTAAACAGCACCAATTGGCGCCCTTGCGTGTGGTG 1013

OY 71 tggtaagttcacatgtaagtgatcttcaatttccacacac 122  
Db 1012 CGTTAAATTCCTGTTAATTGCTAAATATTAGATATTGTTCCACACG 961

## RESULT 3

US-08-477-451-25

; Sequence 25, Application US/08477451  
; Patent No. 5928865  
; GENERAL INFORMATION:  
; APPLICANT: Covacci, Antonello  
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,451  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0335.002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-601-2708  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19932 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-477-451-25

Query Match 20.8%; Score 25.6; DB 2; Length 19932;  
Best Local Similarity 51.8%; Pred. No. 8.9;  
Matches 58; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 11 aaatgtagatgaccttacaacctgatacctaataatgagcgaatgagggctgcaactgac 70  
Db 8085 AAATGCTTGCAATCTGACTATTAAACAGCACCAATTGGCGCCCTTGCGTGTGGTGAG 8144

OY 71 tggtaagttcacatgtaagtgatcttcaatttccacacac 122  
Db 8145 CGTTAAATTCCTGTTAATTGCTAAATATTAGATATTGTTCCACACG 8196

## RESULT 4

US-08-390-878-18

; Sequence 18, Application US/08390878  
; Patent No. 5700683  
; GENERAL INFORMATION:  
; APPLICANT: Stover, Charles K.  
; APPLICANT: Mahairas, Gregory G.  
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: One Market Plaza, Stewart Street Tower, 20th  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/390,878  
; FILING DATE: 17-FEB-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 15371A-17  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/543/9600  
; TELEFAX: 415/543/5043  
; INFORMATION FOR SEQ ID NO: 18:



Query Match	20.38;	Score 25;	DB 2;	Length 2255;
Best Local Similarity	52.48;	Pred. No. 6.5;		
Matches	55;	Conservative	0;	Mismatches 50;
			Indels	0;
			Gaps	0;

Query Match	20.2%	Score 24.8	DB 1	Length 2064
Best Local Similarity	57.9%	Pred. No. 7.4		
Matches 44	Conservative 0	Mismatches 32	Indels 0	Gaps 0
8 tttataatgtgagatgaccttacaacccctatgtacctaatgagcgcaatgcgagagctgcact 67				



Db	633	TGTCTAATGTACGACGAGATGATCACTGGCATGTGTAACCTCAACGGTGACAGGGGTGTCT	6922
QY	68	gactcgttaagtttcaac	83
Db	693	GACTGCTCAGTGGCAC	708

```

RESULT 7
US-08-164-839-32
Sequence 32, Application US/08164839
Patent No. 3514573
GENERAL INFORMATION:
APPLICANT: YASUEDA, HISASHI
APPLICANT: YAKUNISHI, KAZUO
APPLICANT: MOTOKI, MASAO
APPLICANT: NAGASE, KAZUO
APPLICANT: MATSUI, HIROSHI
TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
FROM FISH
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSER: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
City: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/164,839
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/004,729
FILING DATE: 14-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5514573man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-599-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)412-3000
TELEFAX: (703)413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 2064 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Parachanna olivaceus
TISSUE TYPE: liver
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2061
US-08-164-839-32

```

Query Match	20.2%	Score 24.8;	DB 1;	Length 2064;
Best Local Similarity	57.9%;	Pred. No. 7.4;		
Matches 44;	Conservative 0;	Mismatches 32;	Indels 0;	Gaps 0;

[illegible]

Db 693 GACTGGTCACTGGCAC 708

```

US-08-583-799-30      RESULT      8
; Sequence 30, Application US/08583799
; Patent No. 5607849
; GENERAL INFORMATION:
; APPLICANT: YASUEBA, HISASHI
; APPLICANT: NAKANISHI, KAZUO
; APPLICANT: MOTOKI, MASAO
; APPLICANT: NAGASE, KAZUO
; APPLICANT: MATSUI, HIROSHI
; TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
; FROM FISH
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; City: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/583,799
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/004,729
; FILING DATE: 14-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NO. 5607849man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-599-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)412-3000
; TELEFAX: (703)413-2220
; TELEX: 248655 OPAT UR
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2064 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Paratrichthys olivaceus
; TISSUE TYPE: liver
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2061
; US-08-583-799-30

```

Query Match	20.28;	Score 24.8;	DB 1;	Length 2064;
Best Local Similarity	57.98;	Pred. No. 7.4;		
Matches 44;	Conservative	0;	Mismatches 32;	Indels 0;
			Gaps	0;

OY 8 tgaataatgtgagatgcctcaaaccttgatgacctaattggcgaatgcgaggctcact 67  
||| ||||| | | | | | | | |  
Db 633 TGCTATGTCACGAGGATGCATCCTGGGATGGTGAACTTAACCGTGACAGGGGTGCTT 692

```
QY      68 gactgtaagtttcac 83
        |||||  |||  |||
Db      693 GACTGGTCAGTGGCAC 708
```



```

RESULT 9
US-08-583-799-32
: Sequence 32 Application US/08583799
: Patent No. 5607849
: GENERAL INFORMATION:
: APPLICANT: YASUEDA, HISASHI
: APPLICANT: NAKANISHI, KAZUO
: APPLICANT: MOTOKI, MASAO
: APPLICANT: NAGASE, KAZUO
: APPLICANT: MATSUI, HIROSHI
: TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
: FROM FISH
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
: P.C.
: STREET: 1755 Jefferson Davis Highway, Fourth Floor
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/583,799
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/004,729
: FILING DATE: 14-JAN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Oblon, No. 5607849man F.
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 10-599-0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)412-3000
: TELEFAX: (703)413-2220
: TELEX: 248855 OPAT UR
: INFORMATION FOR SEQ ID NO: 32:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2064 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ORIGINAL SOURCE:
: ORGANISM: Paralicthys olivaceus
: TISSUE TYPE: liver
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..2061
: US-08-583-799-32

Query Match 20.2%, Score 24.8, DB 1, Length 2064;
Best Local Similarity 57.9%; Pred. No. 7.4;
Matches 44; Conservativeness 0; Mismatches 32; Indels 0; Gaps 0.0;

OY 8 ttttaattgttgatcccttacaaccttgatgaccctcaatggcgcgaatgagggcgtgaact 67
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 63 ttttctattctcacacagatgatcatctgcgattgtgtgaactcttaacgggtgacaggggttgcct 692
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 68 gactgtgtaagtccac 83
| | | | | | | | | | | |
Db 693 gactgtcactgacgtgcac 708

```

```

: Sequence 69 Application US/08164839
: Patent No. 5514573
:
GENERAL INFORMATION:
APPLICANT: YASUEDA, HISASHI
APPLICANT: NAKANISHI, KAZUO
APPLICANT: MOTOKI, MASAO
APPLICANT: NAGASE, KAZUO
APPLICANT: MATSUI, HIROSHI
TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
FROM FISH
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
ADDRESS: P. C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/164,839
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/004,729
FILING DATE: 14-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5514573man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-599-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)412-3000
TELEFAX: (703)413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 2148 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Parailchthys olivaceus
TISSUE TYPE: liver
FEATURE:
NAME/KEY: CDS
LOCATION: 26..2092
US-08-164-839-69

```



```

: APPLICANT: YASUEDA, HISASHI
: APPLICANT: NAKANISHI, KAZUO
: APPLICANT: MOTOKI, MASAO
: APPLICANT: NAGASE, KAZUO
: APPLICANT: MATSUI, HIROSHI
: TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
: STREET: 1755 Jefferson Davis Highway, Fourth Floor
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/164,839
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/004,729
: FILING DATE: 14-JAN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Oblon, No. 5514573man F.
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 10-599-0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)412-3000
: TELEFAX: (703)413-2220
: TELEX: 248855 OPAT UR
: INFORMATION FOR SEQ ID NO: 71:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2148 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ORIGINAL SOURCE:
: ORGANISM: Paralicthys olivaceus
: TISSUE TYPE: liver
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 26..2092
: US-08-164-839-71

Query Match          20.2%; Score 24.8; DB 1; Length 2148;
Best Local Similarity 57.9%; Pred. No. 7.6;
Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 8 tgaataatgagatgccttacacccctgatacctaattgagcgcaatgcaaggctgcact 67
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 tgcTATGTGTCAGAGATGATCAGTGCATGTGTAACCTTAACGATGAGAGGGGTGTCT 720
QY 68 gactgtaagtttcac 83
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 GACTGTGTCAGTGCAC 736

RESULT 12
US-08-583-799-69
: Sequence 69, Application US/08583799
: Patent No. 5607849
: GENERAL INFORMATION:
: APPLICANT: YASUEDA, HISASHI
: APPLICANT: NAKANISHI, KAZUO
: APPLICANT: MOTOKI, MASAO
: APPLICANT: MATSUI, HIROSHI
```

```

: APPLICANT: NAGASE, KAZUO
: APPLICANT: MATSUI, HIROSHI
: TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: P.C.
: STREET: 1755 Jefferson Davis Highway, Fourth Floor
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/583,799
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/004,729
: FILING DATE: 14-JAN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Oblon, No. 5607849man F.
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 10-599-0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)412-3000
: TELEFAX: (703)413-2220
: TELEX: 248855 OPAT UR
: INFORMATION FOR SEQ ID NO: 69:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2148 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ORIGINAL SOURCE:
: ORGANISM: Paralicthys olivaceus
: TISSUE TYPE: liver
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 26..2092
: US-08-583-799-69

Query Match          20.2%; Score 24.8; DB 1; Length 2148;
Best Local Similarity 57.9%; Pred. No. 7.6;
Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 8 tgaataatgagatgccttacacccctgatacctaattgagcgcaatgcaaggctgcact 67
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 tgcTATGTGTCAGAGATGATCAGTGCATGTGTAACCTTAACGATGAGAGGGGTGTCT 720
QY 68 gactgtaagtttcac 83
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 GACTGTGTCAGTGCAC 736

RESULT 13
US-08-583-799-71
: Sequence 71, Application US/08583799
: Patent No. 5607849
: GENERAL INFORMATION:
: APPLICANT: YASUEDA, HISASHI
: APPLICANT: NAKANISHI, KAZUO
: APPLICANT: MOTOKI, MASAO
: APPLICANT: MATSUI, HIROSHI
: TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
```



```
; TITLE OF INVENTION: FROM FISH
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/583,799
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/004,729
; FILING DATE: 14-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, No. 5607849man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-599-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)412-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2148 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Paralichthys olivaceus
; TISSUE TYPE: liver
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 26..2092
; US-08-583-799-71.

Query Match 20.2%; Score 24.8; DB 1; Length 2148;
Best Local Similarity 57.9%; Pred. No. 7.6;
Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 8 tgaatgtgagatgccttaaaccttgatgacctaatgagcggaatgcgaggtcgtcact 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 TGCTATGTCAGAGGATGATCTACGCGATGTAAGCTTAACGGTACAGGGGTGTGCT 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 68 gactgtgaatttcac 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 GACTGCTCAGTGCAC 736

RESULT 14
US-09-318-448-5/C
; Sequence 5, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.0
```

```
; SEQ ID NO 5
; LENGTH: 7720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-5

Query Match 19.8%; Score 24.4; DB 4; Length 7720;
Best Local Similarity 60.6%; Pred. No. 17;
Matches 40; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 2 aggtatgtgaatgtgagatgcttaaaccttgatgacctaatgagcggaatgcgagggc 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6888 AGGCTGTGTACCTCGAGGAAGCATGATACCTGCTGATCCAGACGGCCCAATTGGGGC 6829
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 tgcact 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6828 AGCACT 6823

RESULT 15
US-08-152-019A-44
; Sequence 44, Application US/08152019A
; Patent No. 5565331
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: Serafini, Tito
; APPLICANT: Kennedy, Timothy
; APPLICANT: Placzek, Marysia
; APPLICANT: Jessell, Thomas
; APPLICANT: Dodd, Jane
; TITLE OF INVENTION: NEURAL AXON OUTGROWTH MODULATORS
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,019A
; FILING DATE: 12-NOV-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard Aron
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59012/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-152-019A-44

Query Match 19.7%; Score 24.2; DB 1; Length 1839;
Best Local Similarity 52.5%; Pred. No. 12;
Matches 53; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 1 caggtatgtgaatgtgagatgcttaaaccttgatgacctaatgagcggaatgcgaggg 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



Db 176 CTGTTATTGCTAGTATCATCGCGGAAACCAACCCAGACAACTATTGTACTGTGAAG 235

Qy 61 ctgcactgacgcgtcaagtttcacatgtaagtgtaacttat 101

Db 236 AGGGTCCGGATGGAATTATTCGGTAGCAGATGTGACACTTGT 276

```

RESULT 16
US-08-482-677-1
Sequence 1, Application US/08482677
Patent No. 6017714
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: Tersifini, Tito
APPLICANT: Kennedy, Timothy
APPLICANT: Placzek, Marysia
APPLICANT: Jessel, Thomas
APPLICANT: Dodd, Jane
TITLE OF INVENTION: Netlins
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,677
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC93-300-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1839 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-482-677-1

```

Query Match	19.7%	Score 24.2	DB 3	Length 1839
Best Local Similarity	52.5%	Pred. No. 12		
Matches 53; Conservative	0	Mismatches 48	Indels 0	Gaps 0

QY 61 ctgcactgactgtgaattccacatgtlaagtgatacttat 101  
| | | | | | | | | | | | | | | | | | | | | |  
Db 236 AGGGTCCGGATGGAAATTTATCCGTGAGCAATGTGACACTTGT 276

RESULT 17  
US-08-727-034-5  
; Sequence 5, Application US/08727034  
; Patent No. 5655872  
; GENERAL INFORMATION:  
; APPLICANT: SATTO, YASHUSHI  
; APPLICANT: IMASAKI, AKIO

APPLICANT: ARAI, KOICHI  
 APPLICANT: YAMAZAKI, HIROYUKI  
 TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND  
 TITLE OF INVENTION: THE GENE CODING THEREFOR  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ADDRESSEE: P.C.  
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 CITY: ARLINGTON  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/727,034  
 FILING DATE: 08-OCT-1996  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 261440/1995  
 FILING DATE: 09-OCT-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 102451/1996  
 FILING DATE: 24-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: OBLON, NORMAN F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 80-079-0  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-413-3000  
 TELEFAX: 703-413-2220  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6642 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 IS-08-727-034-5

Query Match	19.7%	Score 24.2;	DB 1;	Length 6642;
Best Local Similarity	52.5%;	Pred. No. 19;		
Matches 53; Conservative	0;	Mismatches 48;	Indels 0;	Gaps 0;

QY 63 gcaactgactgtaagttcacatgtaagtgatactatct 103  
||| | |||| | ||||| |||||  
Db 5248 CCACCAGATATCCACATGTGACAGCTAGGTGAAATATATCT 5288

RESULT 18  
 US-08-727-034-6  
 ; Sequence 6, Application US/08727034  
 ; Patent No. 5665872  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SAITO, YASHUSHI  
 ; APPLICANT: IMASAKI, AKIO  
 ; APPLICANT: ARAI, KOICHI  
 ; APPLICANT: YAMAZAKI, HIROYUKI  
 ; TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND  
 ; TITLE OF INVENTION: THE GENE CODING THEREFOR  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIYAK, MCLELLAND, MAIER & NEUSADT,  
 ; ADDRESSEE: P.C.



```

: STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
: CITY: ARLINGTON
: STATE: VA
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/727,034
: FILING DATE: 08-OCT-1996
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 261440/1995
: FILING DATE: 09-OCT-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 102451/1996
: FILING DATE: 24-APR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: OBLON, NORMAN F.
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 80-079-0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-413-3000
: TELEFAX: 703-413-2220
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6843 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 81..6725
: OTHER INFORMATION: /note= "Identification Method: S"
: NAME/KEY: sig_peptide
: LOCATION: 81..164
: OTHER INFORMATION: /note= "Identification Method: S"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 165..6722
: OTHER INFORMATION: /function= "Nucleotides 165-6722
: encode the mature peptide"
: OTHER INFORMATION: /note= "Identification Method: S"
: US-08-727-034-6

```

```

Query Match 19.7%; Score 24.2; DB 1; Length 6843;
Best Local Similarity 52.5%; Pred. No. 19;
Matches 53; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
Oy 3 ggatgtgaagtgaagtccttacaccccgatgacctaagggcgcaatggagggct 62
Db 5268 GGAATGGAAGTGGAGCGCATCTTAATCCATTACCAACCAAGAAAGATGCCA 5327
Oy 63 gacatgctgtaagtttcacatgtlaagtgactatct 103
Db 5328 CCACGAGATATCCACATTCAGACGCTATGCTGAATAATTATCT 5368

```

```

RESULT 19
US-08-253-785-2/c
; Sequence 2, Application US/08253785
; Patent No. 5633363
; GENERAL INFORMATION:
; APPLICANT: Colbert, James T.
; APPLICANT: Held, Bruce M.
; APPLICANT: Wurtel, Eve S.
; APPLICANT: Dietrich, Paul S.

```

```

: TITLE OF INVENTION: ROOT PREFERENTIAL PROMOTER
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sandoz Agro, Inc.
: STREET: 975 California Avenue
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/253,785
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Marcus-Wyner, Lynn
: REGISTRATION NUMBER: 34,869
: REFERENCE/DOCKET NUMBER: 135-1089
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/354-3588
: TELEFAX: 415/857-1125
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2329 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 51..1850
: OTHER INFORMATION: /codon_start= 51
: US-08-253-785-2

```

```

Query Match 19.2%; Score 23.6; DB 1; Length 2329;
Best Local Similarity 56.4%; Pred. No. 21;
Matches 44; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
Oy 10 taatgtgagatgctcttacacccctgacctaagcgcaatggcgagctgcactga 69
Db 1001 TAAATCTTCATGCTTGGAGACCATATGACATTTGGCTTATTTGATCATACAC 942
Oy 70 ctgtaagttcacatgt 87
Db 941 ATGGACCATATGATGT 924

```

```

RESULT 20
US-08-975-316-53/c
; Sequence 53, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSBERG, Leonard N., HAYUKKALA, Ilkka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

```



Query Match	19.0%	Score 23.4	DB 2	Length 562
Best Local Similarity	57.5%	Pred. No. 15		
Matches 42	Conservative 0	Mismatches 31	Indels 0	Gaps 0
QY 16	tgaagtccttcaacccttgatgacctaatggcgcaatgagggctgacacgtcgtgta	75		
Db 527	TGATGTCATGCGCTGAAATCTTTCCATCATCGGCAATGCAAGGCGTGTGTGAGACGCG	468		
OY 76	agttcacatgta	88		
Db 467	AGTAACCAAGTGA	455		

Query Match	19.0%	Score 23.4	DB 2	Length 1026
Best Local Similarity	57.5%	Pred. NO. 19		
Matches 42	Conservative	0	Mismatches 31	Indels 0
			Gaps	0
QY	16	tgagatgccttaacccgfatgcactaatggcgaatgcagggcgtccatgacgtga	75	
Db	450	TGATGTCATGCTTAACATCTTTCATCATCTCGGCAATGCAAGGCTGTGCTGAAGAGCG	391	
QY	76	agttcacatgta	88	
Db	390	AGTAACCACTGTA	378	

```

US-08-975-316-6/c
; Sequence 6, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; NUMBER OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
; CORRESPONDENCE ADDRESS: 88
; ADDRESS: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,316
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/713,000
; FILING DATE: September 11, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SLEATH, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000/1003C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: Linear
;
US-08-975-316-6

```



	Query Match	18.7%	Score 23	DB 4	Length 1959
	Best Local Similarity	74.4%	Pred. No. 33		
	Matches 29	Conservative 0	Mismatches 10	Indels 0	Gaps 0
Qy	10	taaatgtgaagatgccttaacaacctgatacctaattgac	48		
Ob	1005	tcaatatgataagacccttaataacctcgggtgataccaattgac	1043		

	Query Match	18.7%	Score 23	DB 1	Length 2059
	Best Local Similarity	55.7%	Pred. No. 34		
	Matches 44	Conservative 0	Mismatches 35	Indels 0	Gaps 0
OY	43	aatggcgaatagcgagggctgcacgcacgtgrraaatttcacatgtaaggtgtaactatcc	102		
Db	1322	ATTGGAGACATCGGGGGTCTCGCTCTCTACGCTCTTAAGGTAGGCGCAATTAAAGGATATTATTC	1381		
OY	103	ttcaattatcttcacacaa	121		



Db 1382 CTCGAGATATCTATACCTA 1400

## RESULT 26

US-08-369-796-7/C  
 ; Sequence 7, Application US/08369796  
 ; Patent No. 5716622  
 ; GENERAL INFORMATION:  
 ; APPLICANT: James E. Darnell, Jr.  
 ; APPLICANT: Zilong Wen  
 ; APPLICANT: Curt M. Horvath  
 ; APPLICANT: Zhong Zhong  
 ; TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL  
 ; TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Klauber & Jackson  
 ; STREET: 411 Hackensack Avenue  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07601  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/369,796  
 ; FILING DATE: 06-JAN-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson Esq., David A.  
 ; REGISTRATION NUMBER: 26,742  
 ; REFERENCE/DOCKET NUMBER: 600-1-116  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201 487-5800  
 ; TELEFAX: 201 343-1684  
 ; TELEX: 133521  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2277 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: both  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: CDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Mouse  
 ; IMMEDIATE SOURCE:  
 ; CLONE: Murine Stat91  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 5..2251  
 ; US-08-369-796-7

Query Match 18.7%; Score 23; DB 1; Length 2277;  
 Best Local Similarity 68.1%; Pred. No. 35;  
 Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 70 ctgtaagtttcacatgtaagtgatacttacttcaattatcca 116  
 ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |  
 Db 538 CTGAGAGGTTTGCAATTAAAGTCATATTCATCTGTGTAATCTCTCA 492

RESULT 27  
 US-08-852-091-7/C  
 ; Sequence 7, Application US/08852091  
 ; Patent No. 5883228  
 ; GENERAL INFORMATION:

APPLICANT: James E. Darnell, Jr.  
 ; APPLICANT: Zilong Wen  
 ; APPLICANT: Curt M. Horvath  
 ; APPLICANT: Zhong Zhong  
 ; TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL  
 ; TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Klauber & Jackson  
 ; STREET: 411 Hackensack Avenue  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07601  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/852,091  
 ; FILING DATE: 06-MAY-1997  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/369,796  
 ; FILING DATE: 06-JAN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson Esq., David A.  
 ; REGISTRATION NUMBER: 26,742  
 ; REFERENCE/DOCKET NUMBER: 600-1-116  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201 487-5800  
 ; TELEFAX: 201 343-1684  
 ; TELEX: 133521  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2277 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: both  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: CDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Mouse  
 ; IMMEDIATE SOURCE:  
 ; CLONE: Murine Stat91  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 5..2251  
 ; US-08-852-091-7

Query Match 18.7%; Score 23; DB 2; Length 2277;  
 Best Local Similarity 68.1%; Pred. No. 35;  
 Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 70 ctgtaagtttcacatgtaagtgatacttacttcaattatcca 116  
 ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |  
 Db 538 CTGAGAGGTTTGCAATTAAAGTCATATTCATCTGTGTAATCTCTCA 492

RESULT 28  
 US-08-820-754-7/C  
 ; Sequence 7, Application US/08820754  
 ; Patent No. 5976835  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Darnell Jr., James E.  
 ; APPLICANT: Schindler, Christian W.  
 ; APPLICANT: Fu, Xian-Yuan  
 ; APPLICANT: Wen, Zilong  
 ; APPLICANT: Zhong, Zhong  
 ; TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN



```

APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey.
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,652
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mouse
IMMEDIATE SOURCE:
CLONE: Murine Stat91
FEATURE:
NAME/KEY: CDS
LOCATION: 5..2251
US-08-956-652-7
Query Match 18.7%; Score 23; DB 3; Length 2277;
Best Local Similarity 68.1%; Pred. No. 35;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
70 ctgtagattcacatgtgaagtgatgaactatcttcaattatcca 116
||||| ||||| ||| ||| ||| ||||| ||||| ||||| |||
538 CTGAGAGGTTTGCAITTTAAGTCATATTCCTTGTAATTCCTTCTA 492

```



RESULT 30  
US-08-956-869-7/c  
Sequence 7, Application US/08956869  
Patent No. 6030808  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956, 869  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/212,185  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Mouse  
IMMEDIATE SOURCE:  
CLONE: Murine Stat91  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5..2251  
US-08-956-869-7

Query Match 18.7%; Score 23; DB 3; Length 2277;  
Best Local Similarity 68.1%; Pred. No. 35;  
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 70 ctgtaagttcacatgtaagtgatcttacttcaatttca 116  
Db 538 CTGAGAGTTTGGCAATTAAAGTCATATTCATCTGTGTAATCTCTCTA 492  
RESULT 31  
US-08-948-547-7/c  
Sequence 7, Application US/08948547  
Patent No. 6124118  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,547  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/212,185  
FILING DATE: 11-MAR-1994  
APPLICATION NUMBER: US 07/980,498  
FILING DATE: 23-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Mouse  
IMMEDIATE SOURCE:  
CLONE: Murine Stat91  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5..2251  
US-08-948-547-7



QY	7	ttgtaaatgltgagatgcgcttaccacccctgatagcctaataatgagcgaatgcgagagctgcac	66
Db	2910	ttgaaatgactctactactatataccctctcctacacttaaggcctctctctctgttgcgaactc	2969
QY	67	tga	69
Db	2970	tga	2972



```
RESULT 35
US-07-792-600-1/c
; Sequence 1, Application US/07792600
; Patent No. 6008045
; GENERAL INFORMATION:
; APPLICANT: COPELAND, WILLIAM C.
; APPLICANT: WANG, TERESA S.-F.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; NUMBER OF INVENTION: TEMPLATE-DEPENDENT ENZYMATIC SYNTHESIS OF NUCLEIC ACID
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter G. Carroll
; STREET: 220 Montgomery Street, Suite 710
; CITY: San Francisco
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,600
; FILING DATE: 19911115
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: STDU-00097
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4440 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-792-600-1

Query Match 18.7%; Score 23; DB 3; Length 4440;
Best Local Similarity 68.1%; Pred. No. 45;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtagatgccttaacaacctgatagactaatgagcg 49
    ||| |||| | | | | | | | | | | | | | | | | | | | | |
Db 408 GGTATTGTAACACTGCGAGCTTCTTACATTCCCTGTGCTTTATTGCG 362

RESULT 36
US-09-157-021-1/c
; Sequence 1, Application US/09157021A
; Patent No. 6100023
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Design Assay
; FILE REFERENCE: STDU-03484
; CURRENT APPLICATION NUMBER: US/09/157,021A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-157-021-1
```

```
Query Match 18.7%; Score 23; DB 3; Length 4440;
Best Local Similarity 68.1%; Pred. No. 45;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtagatgccttaacaacctgatagactaatgagcg 49
    ||| |||| | | | | | | | | | | | | | | | | | | | | |
Db 408 GGTATTGTAACACTGCGAGCTTCTTACATTCCCTGTGCTTTATTGCG 362

RESULT 37
US-09-156-842-1/c
; Sequence 1, Application US/09156842A
; Patent No. 6103473
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Screening
; FILE REFERENCE: STDU-03485
; CURRENT APPLICATION NUMBER: US/09/156,842A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-156-842-1

Query Match 18.7%; Score 23; DB 3; Length 4440;
Best Local Similarity 68.1%; Pred. No. 45;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtagatgccttaacaacctgatagactaatgagcg 49
    ||| |||| | | | | | | | | | | | | | | | | | | | | |
Db 408 GGTATTGTAACACTGCGAGCTTCTTACATTCCCTGTGCTTTATTGCG 362

RESULT 38
US-09-157-021-35/c
; Sequence 35, Application US/09157021A
; Patent No. 6100023
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Design Assay
; FILE REFERENCE: STDU-03484
; CURRENT APPLICATION NUMBER: US/09/157,021A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 35
; LENGTH: 5433
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-157-021-35

Query Match 18.7%; Score 23; DB 3; Length 5433;
Best Local Similarity 68.1%; Pred. No. 49;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtagatgccttaacaacctgatagactaatgagcg 49
    ||| |||| | | | | | | | | | | | | | | | | | | | | |
Db 408 GGTATTGTAACACTGCGAGCTTCTTACATTCCCTGTGCTTTATTGCG 362
```











```

? APPLICANT: KORRKO, JARMO
? APPLICANT: ALA-KORRKO, LEENA, et al.
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
? TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
? NUMBER OF SEQUENCES: 666
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
? STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
? CITY: FLR
? STATE: PHILADELPHIA
? COUNTRY: USA
? ZIP: 19103-7086
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/943,711
? FILING DATE: 03-OCT-1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/212,322
? FILING DATE: 14-MAR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/803,628
? FILING DATE: 03-DEC-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: DOYLE LEARY Ph. D., KATHRYN
? REGISTRATION NUMBER: 36,317
? REFERENCE/DOCKET NUMBER: 9598-27
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-965-1284
? TELEFAX: 215-567-2991
? TELEX: 831-494
? INFORMATION FOR SEQ ID NO: 81:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 911 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-943-731-81

Query Match 18.5%; Score 22.8; DB 4; Length 911;
Best Local Similarity 66.0%; Pred. No. 30;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 64 cactgactgtaagttaacatgtaagtgatctatctcaattatt 113
DB 774 CACTAGCTGGGAAGTGTGAAGAGATTCTGAAATTTCTCTGATTTT 725

RESULT 45
US-08-884-324-2/c
? Sequence 2, Application US/08884324
? Patent No. 6060283
? GENERAL INFORMATION:
? APPLICANT: Takanori OKURA
? APPLICANT: Kakuji TORIGOE
? APPLICANT: Masahi KURIMOTO
? TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
? NUMBER OF SEQUENCES: 35
? TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: BROWDY AND NEIMARK
? STREET: 419 Seventh Street, N.W., Suite 300
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20004
```

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent in Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/884,324
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 185,305/96
? FILING DATE: 27-JUN-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: BROWDY, Roger L.
? REGISTRATION NUMBER: 25,618
? REFERENCE/DOCKET NUMBER: OKURA-1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-628-5197
? TELEFAX: 202-737-3528
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1120 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? MOLECULE TYPE: cDNA to mRNA
? TOPOLOGY: linear
? HYPOHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: human
? TISSUE TYPE: liver
? FEATURE:
? NAME/KEY: 5'UTR
? LOCATION: 1..177
? IDENTIFICATION METHOD: E
? NAME/KEY: leader peptide
? LOCATION: 178..285
? IDENTIFICATION METHOD: S
? NAME/KEY: mat peptide
? LOCATION: 286..756
? IDENTIFICATION METHOD: S
? NAME/KEY: 3'UTR
? LOCATION: 757..1120
? IDENTIFICATION METHOD: E
? US-08-884-324-2

Query Match 18.5%; Score 22.8; DB 3; Length 1120;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 57; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 3 gataatgtaaatgtagagatcccttaacaacccatgacctaattgacgaaggct 62
DB 337 GGTCATTCATCAATTTCTATGACGTGATTAATTGATTCAGCTGCCAAGTAATCTGATT 278

QY 63 gcaactgactgtaagttaacatgtaagtgatctatctcaattatt 116
DB 277 CCAGGTTTCATCATCTTCAGCTATTAAGTAAGCGGATTTGCAATAATTGCA 224
```

Search completed: March 19, 2002, 09:58:41  
Job time: 1615 sec



---



OM of: US-09-684-016-48411 to: PIR\_68:\* out\_format : pfs

Date: Mar 19, 2002 9:32 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

#### Command line parameters:

-MODEL=frame+np model -DEV=x1h  
-O=/cgn2.1/USPRO.spool/US09684016/unat\_19032002.084440\_20919/app-query.fasta.1.182  
-DB=PIR\_68 -GFWF=fastan -SUFFIX=std.rpr -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOEXT=0.000  
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=Dlosum62  
-THRS=human40.cdi -LIST=45 -DOCALLGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09684016.ecgn1.1.56 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

#### Search information block:

Query: US-09-684-016-48411

Query length: 123

Database: PIR\_68:\*

Database sequences: 219241

Database length: 76174552

Search time (sec): 37.240000

#### score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
PIR2:C85253	+	138.00	341.52	201	receptor like protein (partial)
PIR2:T49121	+	138.00	341.52	201	receptor like protein - Arabid
PIR2:S44281	+	127.00	304.10	227	ES43 protein - barley
PIR2:T06070	+	124.00	304.10	258	ES43 protein homolog F19H22.200
PIR2:H85462	+	124.00	304.10	258	ES43 like protein [imported]
PIR2:S69047	+	66.50	155.12	353	hypothetical protein YP138c -
PIR2:T40419	+	65.50	153.03	329	mi02 protein - fission yeast (5
PIR2:T41449	+	62.50	143.69	122	424
PIR2:T21435	+	62.00	141.99	145	probable phd finger transcript
PIR2:T21430	+	62.00	141.19	140	hypothetical protein F26H11.3c
PIR2:T03455	+	60.00	121.04	191	hypothetical protein F26H11.3b
PIR2:T03454	+	60.00	120.64	189	ALR protein - human
PIR2:T25101	+	59.50	136.15	3.27	5262
PIR2:T08905	+	59.50	135.74	3.24	443
PIR2:H81911	+	57.00	139.99	4.33	192
PIR2:T38660	+	57.00	120.90	6.05	1588
PIR2:S48729	+	56.50	125.86	8.22	619
PIR2:H72647	+	56.00	125.44	9.88	344
PIR2:T42156	-	55.00	119.73	8.67	1288
PIR2:T26371	+	54.50	114.24	13.64	368
PIR2:T32633	+	53.00	125.09	31.21	1655
PIR2:A31788	+	53.00	124.42	30.74	180
PIR2:P00542	+	52.00	122.04	42.83	199
PIR2:H86073	+	52.00	122.04	42.83	194
PIR2:S40803	+	52.00	120.63	42.83	194
PIR2:H70109	+	52.00	110.15	36.94	1518
PIR2:T41530	+	52.00	113.35	35.10	721
PIR2:T06497	-	52.00	112.59	34.49	809
PIR2:T16487	-	51.50	118.69	48.01	359
PIR2:A40084	+	51.50	117.35	48.66	335
PIR2:T38168	+	51.50	110.15	40.27	859
PIR2:D96660	+	51.00	124.12	66.22	95
PIR2:T26611	+	51.00	119.12	58.94	205
PIR2:F64828	+	51.00	119.12	58.94	205
PIR2:H85615	+	51.00	117.96	57.37	245
PIR2:T21654	+	51.00	114.76	53.34	346
PIR2:T04224	+	51.00	113.79	52.16	459
PIR2:A35291	-	51.00	112.26	50.37	578
PIR2:F71844	-				

PIR2:H86457 + 51.00 111.02 48.96 697 | hypothetical protein AAG2160  
PIR2:T15670 + 50.50 105.46 52.29 1332 | hypothetical protein C27H5  
PIR2:S22412 + 50.00 117.42 83.42 180 | signal peptidase (EC 3.4.99  
PIR2:S61299 - 50.00 113.29 75.89 336 | lipopolysaccharide heptosylt  
PIR2:B81669 - 50.00 113.29 75.89 336 | lipopolysaccharide heptosylt

seq\_name: PIR2:C85253

#### seq\_documentation block:

receptor like protein (partial) [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: C85253  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp  
Nature 402, 769-777, 1999  
A:title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:reference number: A85001; M01D:20083488  
A:Accession: C85253  
A:Status: preliminary  
A:molecule type: DNA  
A:Residues: 1-201 <STO>  
A:Cross-references: GB:NC\_001268; NID:97269059; PIDN:CAB79169.1; GSPDB:GN00140  
C:Genetics:  
A:gene: AT4G22140  
A:Map position: 4

#### alignment\_scores:

Quality: 138.00 Length: 23  
Ratio: 6.273 Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 91.304

#### alignment block:

US-09-684-016-48411 x C85253 ..  
Align seq 1/1 to: C85253 from: 1 to: 201

5 TATGTAAAGTGGATGACCTTACACCGTACGACCTAATGGCGAATG 54  
|||||  
115 TTYCTSYSCYSGIUMETPROTYRASHNPASPSPLEUETVALGINCY 131  
|||||  
55 CGAGCGCTGCACCTGCTGG 73  
|||||  
131 SGUUGLYCYSLYSASPTRP 137

seq\_name: PIR2:T49121

#### seq\_documentation block:

receptor like protein - Arabidopsis thaliana (fragment)  
N:Alternate names: protein AT4G22140  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49121  
R:Bevan, M.; Weiler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K  
submitted to the Protein Sequence Database, May 2000  
A:reference number: Z25016  
A:Accession: T49121  
A:Status: preliminary  
A:molecule type: DNA  
A:Residues: 1-201 <BEV>  
A:Cross-references: EMBL:AL022140; GSPDB:GN00062; ATSP:AT4G22140  
A:Experimental source: cultivar Columbia; BAC clone F1N20  
C:Genetics:  
A:gene: ATSP:AT4G22140  
A:Map position: 4  
A>Note: intron positions not resolved

#### alignment\_scores:

Quality: 138.00 Length: 23  
Ratio: 6.273 Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 91.304



## alignment\_block:

US-09-684-016-48411 x T49121 ..

Align seg 1/1 to: T49121 from: 1 to: 201

```
5 TATTGTAATGTGAGATGCTTACAACTGATGACCTAATGCGCAATG 54
|||||
115 TTYCYSLYSCYSGlUmEtProTYrAsnProAspAspLeuMeTValGlnCY 131
|||||
55 CGAGGGCTGCACCTGACTGG 73
|||||
131 sglUglnCYsSerGluTrp 137
```

seq\_name: p1r2:S44281

## seq\_documentation\_block:

ES43 protein - barley

C:Species: Hordeum vulgare (barley)

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 24-Mar-1999

C:Accession: S44281

R:Speulman, E.; Salami, F.

submitted to the EMBL Data Library, February 1994

A:Reference number: S44281

A:Accession: S44281

A:Molecule type: mRNA

A:Residues: 1-227 &lt;SPE&gt;

A:Cross-references: EMBL:X71575; NID:9482918; PID:q1345528

C:Genetics:

A:Gene: ES43

## alignment\_scores:

Quality:	127.00	Length:	23
Ratio:	6.048	Gaps:	0
Percent Similarity:	91.304	Percent Identity:	82.609

## alignment\_block:

US-09-684-016-48411 x S44281 ..

Align seg 1/1 to: S44281 from: 1 to: 227

```
5 TATTGTAATGTGAGATGCTTACAACTGATGACCTAATGCGCAATG 54
|||||
147 PheCYSLYSCYSGlUmEtProTYrAsnProAspAspLeuMeTValGlnCY 163
|||||
55 CGAGGGCTGCACCTGACTGG 73
|||||
163 sglUglnCYsSerAspTrp 169
```

seq\_name: p1r2:T06070

## seq\_documentation\_block:

ES43 protein homolog F19H22.200 - Arabidopsis thaliana (fragment)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C:Accession: T06070

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15184

A:Accession: T06070

A:Molecule type: DNA

A:Residues: 1-258 &lt;BEV&gt;

A:Cross-references: EMBL:AL035679; GSPDB:GN00062; ATSP:F19H22.200

C:Experimental source: cultivar Columbia; BAC clone F19H22

C:Genetics:

A:Gene: ATSP:F19H22.200

A:Map position: 4

A:Introns: 25/1; 140/2; 163/2

## alignment\_scores:

Quality:	124.00	Length:	23
Ratio:	5.636	Gaps:	0

Percent Similarity: 95.652 Percent Identity: 78.261

## alignment\_block:

US-09-684-016-48411 x T06070 ..

Align seg 1/1 to: T06070 from: 1 to: 258

```
5 TATTGTAATGTGAGATGCTTACAACTGATGACCTAATGCGCAATG 54
|||||
141 PheCYSLYSCYSGlUmEtProTYrAsnProAspAspLeuMeTValGlnCY 157
|||||
55 CGAGGGCTGCACCTGACTGG 73
|||||
157 sglUglnCYsSerGluTrp 163
```

seq\_name: p1r2:H85462

## seq\_documentation\_block:

ES43 like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C:Accession: H85462

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A&gt;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: H85462

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 &lt;STO&gt;

A:Cross-references: GB:NC\_001268; NID:97270893; PID:NCAB80573.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g39100

A:Map position: 4

## alignment\_scores:

Quality:	124.00	Length:	23
Ratio:	5.636	Gaps:	0
Percent Similarity:	95.652	Percent Identity:	78.261

## alignment\_block:

US-09-684-016-48411 x H85462 ..

Align seg 1/1 to: H85462 from: 1 to: 258

```
5 TATTGTAATGTGAGATGCTTACAACTGATGACCTAATGCGCAATG 54
|||||
141 PheCYSLYSCYSGlUmEtProTYrAsnProAspAspLeuMeTValGlnCY 157
|||||
55 CGAGGGCTGCACCTGACTGG 73
|||||
157 sglUglnCYsSerGluTrp 163
```

seq\_name: p1r2:S69047

## seq\_documentation\_block:

hypothetical protein YPL138c - Yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 05-Nov-1999

C:Accession: S69047

R:Hall, J.; DePaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.

submitted to the EMBL Data Library, December 1995

A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A:Reference number: S69040

A:Accession: S69047

A:Molecule type: DNA

A:Residues: 1-353 &lt;HAL&gt;

A:Cross-references: EMBL:U43703; NID:91244769; PID:q1244777; GSPDB:GN00016; MIPS:YPL1

C:Genetics:

A:Gene: MIPS:YPL138c

A:Map position: 16L

## alignment\_scores:

Quality:	124.00	Length:	23
Ratio:	5.636	Gaps:	0



alignment\_scores:  
 Quality: 66.50 Length: 23  
 Ratio: 4.156 Gaps: 1  
 Percent Similarity: 69.565 Percent Identity: 47.826

alignment\_block:

US-09-684-016-48411 x S69047 ..

Align seg 1/1 to: S69047 from: 1 to: 353

```

5 TATTGTAATGTGAGATGCTTACACCTGATGACCTAATGGCGCAATG 54
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
24 TyrcysilleycyslysglnPro...AsptylglyluleuMetValglyCy 39
55 CGAGGCTGCACGACTGCG 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39 saspGlycysaspPsrp 45

```

seq\_name: p1r2:T40419

seq\_documentation\_block:

m102 protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T40419  
 R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Brown, S.; Harris, D.  
 submitted to the EMBL Data Library, October 1999  
 A:Reference number: Z21927  
 A:Accession: T40419  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-329 <MCD>  
 A:Cross-references: EMBL:AL121863; PIDN:CAB58404.1; GSPDB:GN00067; SPDB:SPBC4.05  
 A:Experimental source: strain 972h; cosmid c4  
 C:Genetics:  
 A:Gene: SPDB:SPBC4.05  
 A:Map position: 2

alignment\_scores:

Quality: 65.50 Length: 27  
 Ratio: 4.367 Gaps: 1  
 Percent Similarity: 55.556 Percent Identity: 44.444

alignment\_block:

US-09-684-016-48411 x T40419 ..

Align seg 1/1 to: T40419 from: 1 to: 329

```

2 AGGATGTGAATGTGAGATGCTTACACCT.....GATGACT 42
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 ArpPheCysIleCysAspPsrpValTyrAsnProGluThrGlnGlyTh 137
43 AATGGCGCAATGGAGGCTGACACTGCG 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
137 rMetPheGlnCysIleLeuGlnGlnAspTrp 147

```

seq\_name: p1r2:T41449

seq\_documentation\_block:

Probable phd finger transcription regulator - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 07-Dec-1999  
 C:Accession: T41449  
 R:Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, September 1998  
 A:Reference number: Z21994  
 A:Accession: T41449  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-424 <RIE>  
 A:Cross-references: EMBL:AL031523; PIDN:CAA20664.1; GSPDB:GN00068; SPDB:SPCC594.05c

A:Experimental source: strain 972h; cosmid c594  
 C:Genetics:  
 A:Gene: SPDB:SPCC594.05c  
 A:Map position: 3

alignment\_scores:

Quality: 62.50 Length: 25  
 Ratio: 4.167 Gaps: 2  
 Percent Similarity: 60.000 Percent Identity: 48.000

alignment\_block:

US-09-684-016-48411 x T41449 ..

Align seg 1/1 to: T41449 from: 1 to: 424

```

5 TATTGTAATGTGAGATGCTTACACCTGATGAC.....CTAATGCG 48
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 Tyrcysilleycysgln.....LysProAspAspLysSerTrpMetle 133
49 GCATGGGAGGCTGCACGACTGCG 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133 uGlyCysAspLysGlnAspTrp 141

```

seq\_name: p1r2:T21433

seq\_documentation\_block:

hypothetical protein F26H11.3a - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000  
 C:Accession: T21433  
 R:Barlow, K.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19421  
 A:Accession: T21433  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-405 <WII>  
 A:Cross-references: EMBL:Z81515; PIDN:CAB04198.1; GSPDB:GN00020; CESP:F26H11.3a  
 A:Experimental source: clone F26H11  
 C:Genetics:  
 A:Gene: CESP:F26H11.3a  
 A:Map position: 2  
 A:introns: 11/1; 42/3; 183/2; 243/2; 287/1  
 C:Superfamily: Bromodomain homology  
 F:266-321/Domain: Bromodomain homology <BRO>

alignment\_scores:

Quality: 62.00 Length: 23  
 Ratio: 4.429 Gaps: 0  
 Percent Similarity: 60.870 Percent Identity: 34.783

alignment\_block:

US-09-684-016-48411 x T21433 ..

Align seg 1/1 to: T21433 from: 1 to: 405

```

5 TATTGTAATGTGAGATGCTTACACCTGATGACCTAATGGCGCAATG 54
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 TyrcysValysGlnGlnProTyrAspAspThrLysPheTyrValglyCy 188
55 CGAGGCTGCACGACTGCG 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
188 saspSerCysGlnGlyTrp 194

```

seq\_name: p1r2:T21435

seq\_documentation\_block:

hypothetical protein F26H11.3c - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000  
 C:Accession: T21435







alignment\_block:  
US-09-684-016-48411 x T03454 ..

Align seg 1/1 to: T03454 from: 1 to: 5262

14 TGTGAGATGCTTACAACTGATGACCTATATGGCGCAATCGAGGCTG 63  
||||| ||||| ::|||::: |||||::: ||  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25101  
1235 CysHisAlaProTyrValGluGluAspLeuLeuIleGlnCysArgHisCys 1251

64 CACTGACTCG 73  
|||  
1251 SGLuArgTyr 1254

seq\_name: p1r2:T25101

seq\_documentation\_block:

hypothetical protein T22C1.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T25101

R:McMurray, A.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19982

A:Accession: T25101

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-416 <WILD>

A:Cross-references: EMBL:Z75550; PIDN:CA99920.1; GSPDB:GN00019; CESP:T22C1.1

A:Experimental source: clone T22C1

C:Genetics:

A:Gene: CESP:T22C1.1

A:Map position: 1

A:introns: 60/1; 80/1; 227/3; 256/3; 369/1

alignment\_scores:  
Quality: 59.50 Length: 27  
Ratio: 3.719 Gaps: 1  
Percent Similarity: 59.259 Percent Identity: 37.037

alignment\_block:  
US-09-684-016-48411 x T25101 ..

Align seg 1/1 to: T25101 from: 1 to: 416

2 AGGATGTGAATGTGAGATGCTTACACCTGATGAC.....CT 42  
:::||||| |||::: ||| ::|||

197 LysPheCysThrCysAspValPheTyrProAspGluAspGlyLysG1 213

43 AATGGCGCAATGGAGGCTGCACTGACTCG 73  
::: ||||||| ||| |||||

213 uLeuLeuGlnCysGlnIleCysGluAspTyr 223

seq\_name: p1r2:T08905

seq\_documentation\_block:

hypothetical protein T32A16.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999

C:Accession: T08905

R:Bevan, M.; Zimmermann, W.; Gruenewald, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May

submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16518

A:Accession: T08905

A:Molecule type: DNA

A:Residues: 1-443 <BEV>

A:Cross-references: EMBL:AL078468; GSPDB:GN00062; ATSP:T32A16.30

A:Experimental source: cultivar Columbia; BAC clone T32A16

C:Genetics:

A:Map position: 4

A:introns: 28/3; 81/3; 172/3

alignment\_scores:

Quality: 59.50 Length: 29  
Ratio: 4.250 Gaps: 2  
Percent Similarity: 48.276 Percent Identity: 44.828

alignment\_block:  
US-09-684-016-48411 x T08905 ..

Align seg 1/1 to: T08905 from: 1 to: 443

5 TATGTAAATGTGAGATGCTTACACCTGATGAC..... 40  
||||| |||::: ||||| |||||  
128 TyrCysThrCysAspArgProTyr...ProAspProAsnValGluGlu 143

41 ....CTAATGGCGCAATGGAGGCTGCACTGACTCG 73  
||| ||||| ||| |||||  
143 nValGluMetIleGlnCysCysLeuGluAspTyr 155

seq\_name: p1r2:B81911

seq\_documentation\_block:

probable molybdopterin-guanine dinucleotide biosynthesis protein A MMA1417 [imported]

C:Species: Neisseria meningitidis

C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: B81911

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

; Holtroyd, S.; Jørgensen, K.; Leather, S.; Mould, S.; Mungall, K.; Quail, M.A.; Rajandre

Nature 404, 502-506, 2000

A:title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491

A:Reference number: AB1775; MUID:20222556

A:Accession: B81911

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-192 <PAR>

A:Cross-references: GB:AL162755; GB:AL157959; NID:97379742; PIDN:CAB84657.1; PID:9738

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: MMA1417

alignment\_scores:  
Quality: 59.00 Length: 15  
Ratio: 4.538 Gaps: 0  
Percent Similarity: 86.667 Percent Identity: 66.667

alignment\_block:  
US-09-684-016-48411 x B81911 ..

Align seg 1/1 to: B81911 from: 1 to: 192

14 TGTGAGATGCTTACAACTGATGACCTATATGGCGCAATGGCGAG 58  
||||:||||| ||||| |||||::: |||  
100 CysAspMetProTyrLeuProAspAspLeuValAlaArgPheGlu 114

seq\_name: p1r2:T38660

seq\_documentation\_block:

probable transcription regulator protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T38660

R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z21804

A:Accession: T38660

A:Molecule type: DNA

A:Residues: 1-1588 <MR>

A:Cross-references: EMBL:AL109739; NID:el534774; PIDN:CAB52274.1; GSPDB:GN00066; SPDB

A:Experimental source: strain 972h-; cosmid G343

C:Genetics:

A:Gene: SPDB:SPAC343.11c



A:Map position: 1

alignment\_scores:  
Quality: 57.00 Length: 22  
Ratio: 3.562 Gaps: 0  
Percent Similarity: 72.727 Percent Identity: 36.364

alignment\_block:

US-09-684-016-48411 x T38660 ..  
Align seg 1/1 to: T38660 from: 1 to: 1588

8 TGTAAATGATGATGCTTACAAACCTGATGACCTAATGGCCAAATGCGA 57  
||||| .....  
1457 CysLeuCySArgInProPhaIaIeSerApspLYThrValGIncySHL 1473  
58 GGGCTGCACCTGACTGG 73  
.....  
1473 sasncysleuGIUTTP 1478

seq\_name: p1r1:S48729

seq\_documentation\_block:

glucokinase regulator - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S48729; S49339  
R:Val9-da-Cunha, M.; Deltheux, M.; Watelet, N.; van Schaftingen, E.  
Eur. J. Biochem. 225, 43-51, 1994  
A:Title: Cloning and expression of a Xenopus liver cDNA encoding a fructose-phosphate-in-  
A:Reference number: S48729; MUID:95010134  
A:Accession: S48729  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-619 <VEI>  
A:Cross-references: EMBL:X80901; NID:9556677; PIDN:CAAS6863.1; PID:9556678  
C:Superfamily: glucokinase regulator

alignment\_scores:  
Quality: 56.50 Length: 28  
Ratio: 2.568 Gaps: 1  
Percent Similarity: 78.571 Percent Identity: 39.286

alignment\_block:

US-09-684-016-48411 x S48729 ..  
Align seg 1/1 to: S48729 from: 1 to: 619

3 GGTATTGTAAT...GTGAGATGCTTACACCTGATGACCTAATGGCG 49  
||||| .....  
341 GYIleMetasPalavaiGIncySValProInrTYrGInaIaIasPrIprr 357  
50 CAATCGAGGCGCTGCATGCTGATGCTAAGTTTAC 83  
|:.....: .....  
357 gaspvalarGgIyPheIleThrGlyGlyTyrrHis 368

seq\_name: p1r2:H72647

seq\_documentation\_block:

hypothetical protein APE0614 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: H72647  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum  
A:Reference number: A72450; MUID:99310339  
A:Accession: H72647  
A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-544 <KAW>  
A:Cross-references: DDBJ:AP00060; NID:95104188; PIDN:BAAT79584.1; PID:01043370; PID:9  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0614

alignment\_scores:  
Quality: 56.00 Length: 26  
Ratio: 3.111 Gaps: 0  
Percent Similarity: 69.231 Percent Identity: 38.462

alignment\_block:

US-09-684-016-48411/rev x H72647 ..  
Align seg 1/1 to: H72647 from: 1 to: 544

116 TGGAAATTAATGAAGTAATGATACCTTACATGCTAACTTACAGTCA 67  
||||| .....  
179 TrpGluProIleaspIaValPheIysValGIuIleGluIleuaspArgse 195  
195 rCySerProArgIleAlaIaIaIeucly 204

seq\_name: p1r2:T42756

seq\_documentation\_block:

5-oxoprolinase (ATP-hydrolyzing) (EC 3.5.2.9) - rat  
N:Alternate names: pyroglutamate (ATP-hydrolyzing)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-May-2000  
C:Accession: T42756  
R:Ye, G.J.; Breslow, E.; Meister, A.  
J. Biol. Chem. 271, 32293-32300, 1996  
A:Title: The amino acid sequence of rat kidney 5-oxo-L-prolinase determined by cDNA c  
A:Reference number: Z22263; MUID:97113037  
A:Accession: T42756  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1288 <YEG>  
A:Cross-references: EMBL:U070825; NID:91732064; PID:91732065; PIDN:AAC52955.1  
A:Experimental source: Strain Sprague-Dawley; kidney  
C:Superfamily: hypothetical protein YKL215c  
C:Keywords: homodimer; hydrolase

alignment\_scores:  
Quality: 56.00 Length: 30  
Ratio: 2.800 Gaps: 1  
Percent Similarity: 66.667 Percent Identity: 40.000

alignment\_block:

US-09-684-016-48411/rev x T42756 ..  
Align seg 1/1 to: T42756 from: 1 to: 1288

86 CATGTGAACCTTACAGTACGATGACGCCCT.....CGCATGCCCAT 43  
||||| .....  
215 HisValSerLeuSerSerGluValMetPrometValArgIleValProAr 231  
42 AGGTCAATCAGGGTTGTAGGCAATCTACACTTACATACC 3  
||||| .....  
231 gGIyHISrHraIaCySaIaAsPaIaTyIleuThrProThr 244

seq\_name: p1r2:T26971

seq\_documentation\_block:

hypothetical protein Y47H9C.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26971  
R:Harris, B.











A:Experimental source: strain 97zh-; cosmid c645  
 C:Genetics:  
 A:Gene: SPDB:SPCC645.13  
 A:Map position: 3

alignment\_scores:  
 Quality: 52.00 Length: 22  
 Ratio: 4.333 Gaps: 0  
 Percent Similarity: 54.545 Percent Identity: 36.364

alignment\_block:  
 US-09-684-016-48411 x T41530 ..

Align seg 1/1 to: T41530 from: 1 to: 721

```

8 TGTAAATGTGATGCTTACACCCGTGATGCTTAATGGCGCAATGCGA 57
   ||| |||::: ||| ||| ||| ||| ||| ||| ||| ||| |||
23 CysValCysLysSerGlnGluAspIleGlyAspThrTrpValGlnCysAs 39
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
58 GGGCGGCACTGACTGG 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
39 PclYcysAspCysTrp 44

```

seq\_name: p1r2:T06497

seq\_documentation\_block:

probable sucrose synthase (EC 2.4.1.13) 2 - garden pea

C:Species: Pisum sativum (garden pea)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Jun-2000

C:Accession: T06497

R:Buchner, P.

submitted to the EMBL Data Library, October 1997

A:Reference number: Z15720

A:Accession: T06497

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-809 <BUC>

A:Cross-references: EMBL:AJ001071; PIDN:CAA04512.1

A:Experimental source: cultivar Frisson

C:Superfamily: sucrose synthase; sucrose/sucrose-phosphate synthase homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:282-757/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

alignment\_scores:  
 Quality: 52.00 Length: 57  
 Ratio: 2.364 Gaps: 2  
 Percent Similarity: 38.596 Percent Identity: 26.316

alignment\_block:

US-09-684-016-48411/rev x T06497 ..

Align seg 1/1 to: T06497 from: 1 to: 809

```

105 GAAGATAGATGATCCTTACCTGCGA..... 79
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 GlnAspLysTrpHisPheSerCysGlnPheThrAlaAspLeuIleAlaMe 473
   .....
78 .....
   .....
473 tAsnAsnAlaAspPheIleIleThrSerThrTyrglnGlnIleAlaGlyT 490
   .....
63 .....
   .....
490 hrlYsAsnThrIleGlyGlnTyrglnSerHisThrAlaPheThrLeuPro 506
   .....
33 GGGTTGTAGGCAATGCATCAT 13
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
507 GlyLeuTyraGlyAlaValHis 513

```

seq\_name: p1r2:T16487

seq\_documentation\_block:  
 hypothetical protein F56F10.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 04-Mar-2000  
 C:Accession: T16487  
 R:Bradshaw, H.

submitted to the EMBL Data Library, March 1996  
 A:Description: The sequence of C. elegans cosmid F56F10.  
 A:Reference number: Z18521

A:Accession: T16487

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-269 <BR>

A:Cross-references: EMBL:U51993; NID:g1688050; PID:g1255790; PIDN:AA836851.1; GSPDB:(

A:Experimental source: strain Bristol N2; clone F56F10

C:Genetics:

A:Gene: CESP:F56F10.2

A:Map position: X

A:Introns: 21/3; 45/2; 100/3; 147/3; 177/3; 232/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F56F10.2

alignment\_scores:  
 Quality: 51.50 Length: 37  
 Ratio: 2.060 Gaps: 1  
 Percent Similarity: 67.568 Percent Identity: 32.432

alignment\_block:

US-09-684-016-48411/rev x T16487 ..

Align seg 1/1 to: T16487 from: 1 to: 269

```

119 GTGTGGATTAATGGAAGATA...AGTATCACTTACATGTGAACCTTAC 73
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||
221 IleTrpSerLysPheSerAlaAspSerIlePheLeuPheGlnGluLeuVa 237
   .....
72 CAGTCAGTCAGACCCCTGCGCATTCGCGCATTCAGTCATCAGGCTGTAGG 23
   ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
237 llyGlnAsnSerAspThrValThrAspPheGlyLeuGlnLucySserp 254
   .....
22 CATCTGCATTT 12
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 roSerHisLeu 257

```

seq\_name: p1r2:A40084

seq\_documentation\_block:

exogastrola-inducing protein (EGIP) precursor - sea urchin (Strongylocentrotus purpu

N:Alternate names: epidermal growth factor-related protein

N:Contains: EGIP peptide A; EGIP peptide B; EGIP peptide C; EGIP peptide D

C:Species: Strongylocentrotus purpuratus (purple urchin)

C:Date: 20-Mar-1992 #sequence\_revision 15-Aug-1997 #text\_change 11-Jan-2000

C:Accession: A40084

R:Yang, O.; Angerer, L.M.; Angerer, R.C.

Science 246, 806-808, 1989

A:Title: Unusual pattern of accumulation of mRNA encoding EGF-related protein in sea

A:Reference number: A40084; M0ID:90049203

A:Accession: A40084

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-325 <YAN>

A:Cross-references: GB:M29004; NID:g161463; PID:g161464

A:Note: the publication does not report the complete sequence

C:Superfamily: unassigned EGF-related proteins; EGF homology

F:1-18/Domain: signal sequence #status predicted <SIS>

F:19-325/Product: exogastrola-inducing protein #status predicted <EGIP>

F:47-104/Product: EGIP peptide C #status predicted <EGIC>

F:52-90/Domain: EGF homology <EGI>

F:106-156/Product: EGIP peptide D #status predicted <EGID>

F:111-151/Domain: EGF homology <EG2>

F:177-228/Product: EGIP peptide A #status predicted <EGIA>

F:182-223/Domain: EGF homology <EG3>

F:250-309/Product: EGIP peptide B #status predicted <EGIB>



F:256-282/Domain: EGF homology <EG4>  
F:52-65,59-75,77-90,111-124,118-137,139-151,182-195,189-209,211-223,256-269,263-278,280-

## alignment\_scores:

Quality:	51.50	Length:	14
Ratio:	4.682	Gaps:	1
Percent Similarity:	78.571	Percent Identity:	64.286

## alignment\_block:

US-09-684-016-48411 x A40084 ..

Align seg 1/1 to: A40084 from: 1 to: 325

```

5 TATGTGAATGATGATGCTTACACAC.....CCTGAT 37
|||||
74 TYRCYSGLYSCYSGIUSMETPROPHEARGVALGYLLEPROASP 87

```

seq\_name: p1r2:T38168

## seq\_documentation\_block:

hypothetical protein SPAC22E12.11c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
R:Devlin, K.; Church, C.M.; Barrell, B.G.; Randalream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, March 1996  
A:Accession: T38168  
A:Reference number: 221775  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1859 <DEV>  
A:Cross-references: EMBL:Z70043; NID:G1220275; PIDN:CAA3898.1; GSPDB:GN00066; SPDB:SPAC  
A:Experimental source: strain 972h; cosmid c22E12  
C:Genetics:  
A:Gene: SPDB:SPAC22E12.11c  
A:Map position: 1

## alignment\_scores:

Quality:	51.50	Length:	22
Ratio:	3.679	Gaps:	1
Percent Similarity:	63.636	Percent Identity:	40.909

## alignment\_block:

US-09-684-016-48411 x T38168 ..

Align seg 1/1 to: T38168 from: 1 to: 859

```

11 AATGTGAATGCTTACACCTGATGAC..CTATGGCGCATCGGA 57
|||||
5 ATGCVSVALCYSPROPHGLUASPSAPG1YPHERHRLLEGLNCSGL 21
58 GGGCTGCACCTGACTGG 73
|||||
21 USCYSGIUSVALTRIP 26

```

seq\_name: p1r2:D96660

## seq\_documentation\_block:

protein F2K11.14 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.E.; Hughes, B.; Hultzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.-A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Kerr, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: D96660

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1518 <STO>

A:Cross-references: GB:AE005173; NID:96633837; PIDN:AA19696.1; GSPDB:GN00141

C:Genetics:

A:Gene: F2K11.14

A:Map position: 1

## alignment\_scores:

Quality:	51.50	Length:	23
Ratio:	3.433	Gaps:	1
Percent Similarity:	65.217	Percent Identity:	39.130

## alignment\_block:

US-09-684-016-48411 x D96660 ..

Align seg 1/1 to: D96660 from: 1 to: 1518

```

5 TATGTGAATGATGATGCTTACACCTGATGATGATGATGCGCAATG 54
|||||
1411 HISCYSLLECYSLEULSPROTIFRANSEARISER...MetValsercy 1426
55 CGAGGCTGCACCTGACTGG 73
|||||
1426 SSGGLNCSGLYGLUTRP 1432

```

seq\_name: p1r2:T26611

## seq\_documentation\_block:

hypothetical protein Y37A1B.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
R:McMurray, A.  
submitted to the EMBL Data Library, June 1998  
A:Reference number: Z20245  
A:Accession: T26611  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-95 <ML>  
A:Cross-references: EMBL:AL023835; PIDN:CAA19489.1; GSPDB:GN00022; CESP:Y37A1B.4  
A:Experimental source: clone Y37A1B  
C:Genetics:  
A:Gene: CESP:Y37A1B.4  
A:Map position: 4  
A:introns: 46/2; 84/3

## alignment\_scores:

Quality:	51.00	Length:	36
Ratio:	2.833	Gaps:	1
Percent Similarity:	50.000	Percent Identity:	30.556

## alignment\_block:

US-09-684-016-48411/rev x T26611 ..

Align seg 1/1 to: T26611 from: 1 to: 95

```

122 GTTGTGGAATGAATGAATGAATGATGATGATGATGATGATGATG 73
|||||
46 ILEULTRPSANSLNLEULVALMETILE..... 55
72 CAGTCACTGCAGCCCTGCGCATTTGGCCATTAGGTCAATCAGGGTTGACG 23
|||||
56 ....GlnCysAspGlyAsnIleuLeuProGlnYshGlnGlnYPhetyRA 71
22 CATCTCAC 15
|||||
71 rphphs 73

```



```

seq_name: p1r2:F64828

seq_documentation_block:
  anaerobic dimethylsulfoxide reductase (EC 1.-.-.-) chain B - Escherichia coli
  C:Species: Escherichia coli
  C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 11-Jun-1999
  C:Accession: F64828; S03786; G90169
  R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
    A.; Rose, D.J.; Mau, B.; Shao, Y.
  Science 277, 1453-1462, 1997
  A:Title: The complete genome sequence of Escherichia coli K-12.
  A:Reference number: A64720; MUID:97426617
  A:Accession: F64828
  A:Status: nucleic acid sequence not shown; translation not shown
  A:Molecule type: DNA
  A:Residues: 1-205 <BLAT>
  A:Cross-references: GB:AE000191; GB:U00096; NID:g1787115; PIDN:AACT3981.1; PID:g1787122;
  R:Billions, P.T.; Cole, S.T.; Anderson, W.F.; Weiner, J.H.
  M.O.L. Microbiol. 2, 785-795, 1998
  A:Title: Nucleotide sequence of the dmsABC operon encoding the anaerobic dimethylsulphox
  A:Reference number: S03784; MUID:89096500
  A:Molecule type: DNA
  A:Residues: 1-170,'RA',171-205 <BLT>
  A:Cross-references: GB:U03412; NID:g145754; PIDN:AA83844.1; PID:g145756
  A:Note: part of this sequence, including the amino end of the mature protein, was confir
  C:Genetics:
  A:Gene: dmsB
  A:Map position: 20 min
  C:Complex: heterotrimer: chains A, B, and C
  C:Function:
  A:Description: terminal reductase during anaerobic growth on various sulfoxide and N-oxi
  A:Note: chain A binds molybdopterin, chain B is an electron transfer protein, and chain
  C:Superfamily: nrfC protein; ferredoxin 2[4Fe-4S] homology
  C:Keywords: 4Fe-4S; heterotrimer; iron-sulfur protein; metalloprotein; oxidoreductase
  F:2-205/Product: dimethylsulfoxide reductase chain B #status experimental <MAT>
  F:7-87/Domain: ferredoxin 2[4Fe-4S] homology <FER1>
  F:92-133/Domain: ferredoxin 2[4Fe-4S] homology <FER2>
  F:14,11,20,79/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
  F:24,67,70,75/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
  F:99,102,105,145/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
  F:109,126,129,141/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

alignment_scores:
  Quality: 51.00 Length: 29
  Ratio: 3.400 Gaps: 1
  Percent Similarity: 51.724 Percent Identity: 34.483

alignment_block:
  US-09-684-016-48411 x F64828 ..

Align seg 1/1 to: F64828 from: 1 to: 205

2 AGGATTTGTAATGTGAGATGCTTACCAACCTGATGACCTA..... 43
  |||||
  103 ArgTYrCYSHisMetAlaCysProTYrGlyAlaProGlnTYrAsnGluTh 119
  44 .....ATGGCGCAATGCGAGGCTGCACCTGAC 70
  |||||
  119 rlysglyHisMetThrlyscysaspGlycystyrasp 131

seq_name: p1r2:B85615

seq_documentation_block:
  hypothetical protein dmsB [imported] - Escherichia coli (strain O157:H7)
  C:Species: Escherichia coli
  C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
  C:Accession: B85615
  R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
  11ler, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Potamousis, K.; Apodaca,

```

```

  Nature 409, 529-533, 2001
  A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
  A:Reference number: A85480; MUID:21074935; PMID:11206551
  A:Accession: B85615
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-205 <STO>
  A:Cross-references: GB:AE005174; NID:g12514060; PIDN:AAG55382.1; GSPDB:GN00145; UMG
  A:Experimental source: strain O157:H7, substrain EDL933
  C:Genetics:
  A:Gene: dmsB

alignment_scores:
  Quality: 51.00 Length: 29
  Ratio: 3.400 Gaps: 1
  Percent Similarity: 51.724 Percent Identity: 34.483

alignment_block:
  US-09-684-016-48411 x B85615 ..

Align seg 1/1 to: B85615 from: 1 to: 205

2 AGGATTTGTAATGTGAGATGCTTACCAACCTGATGACCTA..... 43
  |||||
  103 ArgTYrCYSHisMetAlaCysProTYrGlyAlaProGlnTYrAsnGluTh 119
  44 .....ATGGCGCAATGCGAGGCTGCACCTGAC 70
  |||||
  119 rlysglyHisMetThrlyscysaspGlycystyrasp 131

seq_name: p1r2:T21654

seq_documentation_block:
  hypothetical protein F32D8.3 - Caenorhabditis elegans
  C:Species: Caenorhabditis elegans
  C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
  C:Accession: T21654
  R:Wilkinson, J.
  Submitted to the EMBL Data Library, June 1996
  A:Reference number: Z19454
  A:Accession: T21654
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-245 <MTL>
  A:Cross-references: EMBL:Z74031; PIDN:CAA98455.1; GSPDB:GN00023; CESP:F32D8.3
  A:Experimental source: clone F32D8
  C:Genetics:
  A:Gene: CESP:F32D8.3
  A:Map position: 5
  A:Introns: 61/1; 83/3
  C:Superfamily: Caenorhabditis elegans hypothetical protein F32D8.3

alignment_scores:
  Quality: 51.00 Length: 23
  Ratio: 3.643 Gaps: 2
  Percent Similarity: 60.870 Percent Identity: 47.826

alignment_block:
  US-09-684-016-48411 x T21654 ..

Align seg 1/1 to: T21654 from: 1 to: 245

5 TATTGTTAATGTGAGATGCTT...TACAACCTGATGACCTAATGCGCA 51
  ::||| |||||
  71 HisCysAlaCysGluSerThrcysAsnProAspProTYrCysSerly 87
  52 ATGCGAG...GGTCGACT 67
  :||| |||||
  87 scysGluProGlyCysThr 93

seq_name: p1r2:T04224

```



seq\_documentation\_block:  
 hypothetical protein T5C23.200 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 24-Nov-1999  
 C:Accession: T04224  
 R:Bayan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, March 1999  
 A:Reference number: 215261  
 A:Accession: T04224  
 A:Molecule type: DNA  
 A:Residues: 1-396 <BEF>  
 A:Cross-references: EMBL:AL049500  
 A:Experimental source: cultivar Columbia; BAC clone T5C23  
 C:Genetics:  
 A:Map position: 4  
 A:Note: T5C23.200  
 C:Superfamily: caffeoyl-CoA 3-O-methyltransferase

alignment\_scores:  
 Quality: 51.00 Length: 16  
 Ratio: 4.250 Gaps: 0  
 Percent Similarity: 75.000 Percent Identity: 56.250

alignment\_block:  
 US-09-684-016-48411 x T04224 ..  
 Align seg 1/1 to: T04224 from: 1 to: 396

14 TGTGAGATGCGCTTACACCGCTGACCTGATGGCGCAATGCGAGGC 61  
 |||:::|||||||::: |||  
 10 CysAsmMetProTyRLeuProAspAspLeuLeuAsnIleLeuGly 25

seq\_documentation\_block:  
 adenylosuccinate lyase (EC 4.3.2.2) - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 17-Aug-1990 #sequence\_revision 17-Aug-1990 #text\_change 21-Jul-2000  
 C:Accession: A35291  
 R:Almi, J.; Badylak, J.; Williams, J.; Chen, Z.; Zalkin, H.; Dixon, J.E.  
 J. Biol. Chem. 265, 9011-9014, 1990  
 A:Title: Cloning of a cDNA encoding adenylosuccinate lyase by functional complementation  
 A:Reference number: A35291; MUID:90264380  
 A:Accession: A35291  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-459 <AIK>  
 A:Cross-references: GB:M37901; MID:9211094; PIDN:AAA48574.1; PID:9211095  
 C:Superfamily: fumarate hydratase  
 C:Keywords: amidine-lyase; carbon-nitrogen lyase

alignment\_scores:  
 Quality: 51.00 Length: 33  
 Ratio: 2.550 Gaps: 0  
 Percent Similarity: 60.606 Percent Identity: 27.273

alignment\_block:  
 US-09-684-016-48411/rev x A35291 ..  
 Align seg 1/1 to: A35291 from: 1 to: 459

110 AAATTGAAGATAGTATCAGCTTACATGTAACCTTACAGTCAGTCAG 61  
 |||||::: ::::: |||||::: ::::: |||  
 58 LysLeuArgHisAspValMetAlaHisValHisThrPheAlaHisCysCy 74  
 60 CCCTGCGCATTCGCGCATTCAGTCAGTCAGTCAGTCAGTCAGTCATT 12  
 ||::: |||::: ||| ||::: ::::: |||  
 74 SPRLYSAL 90  
 seq\_name: p1r2:F71844

seq\_documentation\_block:  
 ABC transporter, ATP-binding protein - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori  
 A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 17-Mar-2000  
 C:Accession: F71844  
 R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p  
 A:Reference number: A71800; MUID:99120557  
 A:Accession: F71844  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-578 <ARN>  
 A:Cross-references: GB:AE001541; GB:AE001439; MID:94155724; PIDN:AAD06712.1; PID:9415  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: jhp1129  
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog  
 F:381-573/Domain: ATP-binding cassette homolog <ABC>

alignment\_scores:  
 Quality: 51.00 Length: 31  
 Ratio: 2.684 Gaps: 0  
 Percent Similarity: 61.290 Percent Identity: 38.710

alignment\_block:  
 US-09-684-016-48411/rev x F71844 ..  
 Align seg 1/1 to: F71844 from: 1 to: 578

110 AAATTGAAGATAGTATCAGCTTACATGTAACCTTACAGTCAGTCAG 61  
 ||| |||::: ::::: ||||| |||::: |||  
 375 LysSerLysHisProValLeuLysAsnPheAsnLeuThrIleIleInysGl 391  
 : ::|||::: ::|||::: |||  
 60 CCCTGCGCATTCGCGCATTCAGTCAGTCAGTCAGTCAGTCAGTCATT 18  
 : ::|||::: ::|||::: |||  
 391 YGInLysIleAlaLeuIleGlyHisSerGlyCysGlyLysSer 405

seq\_name: p1r2:H86457  
 seq\_documentation\_block:  
 hypothetical protein AAG21600.1 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: H86457  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
 Jensen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mailli, R.; Marzia  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: H86457  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-697 <STO>  
 A:Cross-references: GB:AE005172; MID:910645488; PIDN:AAG21600.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

alignment\_scores:  
 Quality: 51.00 Length: 28  
 Ratio: 3.643 Gaps: 1  
 Percent Similarity: 50.000 Percent Identity: 35.714



## alignment\_block:

US-09-684-016-48411 x H86457 ..

Align seg 1/1 to: H86457 from: 1 to: 697

```

      8 TGTAAATGTGAGATGCTTACACACCCCTGATGACCTTAATGCGCATGCGA 57
      |||||
      606 CyslyscysglYThrlyAspAspAspGlyGluArgMetLeuAlaCysAs 622
      |||||
      58 GCGCTGACAGTGTGTAAGTTTCACATGTAAG 91
      |||||
      622 pGlyCysGlyValTrp.....HisHisThrArg 631

```

seq\_name: p1r2:T15670

## seq\_documentation\_block:

hypothetical protein C27H5.5 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T15670

R:Pauley, A.

submitted to the EMBL Data Library, July 1995

A:Description: The sequence of *C. elegans* cosmid C27H5.

A:Reference number: Z18386

A:Accession: T15670

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1332 &lt;PAU&gt;

A:Cross-references: EMBL:U14635; NID:9540265; PID:9540271; PIDN:MAC46659.1; CESP:C27H5.5

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:C27H5.5

A:Introns: 27/2; 93/3; 203/1; 336/1; 358/2; 394/3; 473/2; 489/3; 531/3; 574/2; 636/3; 67

## alignment\_scores:

Quality:	50.50	Length:	36
Ratio:	2.196	Gaps:	3
Percent Similarity:	63.889	Percent Identity:	36.111

## alignment\_block:

US-09-684-016-48411/rev x T15670 ..

Align seg 1/1 to: T15670 from: 1 to: 1332

```

      95 ATCACTTACATGTGAACCTTACCAAGTCAGTCAGCCCTCCATTGCG.. 48
      ::|||
      176 LeuThrLeuHisLeu.....AlaCysSerMetHisLeuSerAr 188
      |||||
      47 ....CCATTAGTCATCAGGGTTGTAG.....GCATCTGCATT 11
      |||||
      188 gTyrPrProMetAspHisGlnAsnCysGluIleAlaPheAlaSerYrAlaT 205
      10 ACAATACC 3
      |||||
      205 yrThrThr 207

```

seq\_name: p1r2:S22412

## seq\_documentation\_block:

signal peptidase (EC 3.4.99.-), microsomal - chicken

C:Species: *Gallus gallus* (chicken)

C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 21-Jul-2000

C:Accession: S22412; S40018; S15551

R:Newsome, A.L.; McLean, J.W.; Lively, M.O.

Biochem. J. 282, 447-452, 1992

A:title: Molecular cloning of a cDNA encoding the glycoprotein of hen oviduct microsomal

A:Reference number: S22412; MUID:92189580

A:Accession: S22412

A:Molecule type: mRNA

A:Residues: 1-180 &lt;NEWI&gt;

A:Cross-references: EMBL:X60795; NID:963421; PIDN:CAA43208.1; PID:963422

A:Accession: S40018

A:Molecule type: protein

A:Residues: 1-15;36-45;50-58;76-83;97-118;141-160 &lt;NEW2&gt;

C:Keywords: glycoprotein; hydrolase; transmembrane protein

F:1-180/Product: signal peptidase #status experimental &lt;MAT&gt;

F:141/Binding site: carbohydrate (Asn) (covalent) #status experimental

## alignment\_scores:

Quality:	50.00	Length:	28
Ratio:	2.500 <td>Gaps:</td> <td>0 </td>	Gaps:	0
Percent Similarity:	71.429 <td>Percent Identity:</td> <td>39.286</td>	Percent Identity:	39.286

## alignment\_block:

US-09-684-016-48411 x S22412 ..

Align seg 1/1 to: S22412 from: 1 to: 180

```

      38 GACCTAATGCGGCAATGCGAGGCTGCACTGACGTGAATTTCACATGT 87
      |||||
      69 AspIleThrAlaAspLeuGlnSerIlePheAspTrpAsnValylsGlnIle 85
      |||||
      88 AAGGTGATCTTATCTTCATTTATTCACACAA 121
      |||||
      85 upheutyrleuSerAlaGluTrpSerThrlys 96

```

seq\_name: p1r2:S61299

## seq\_documentation\_block:

lipopolysaccharide heptosyltransferase (EC 2.4.99.-) II rfaI [similarity] - *Neisseri*C:Species: *Neisseria gonorrhoeae*

C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Jul-2000

C:Accession: S61299

R:Schwan, E.T.; Robertson, B.D.; Brade, H.; van Putten, J.P.M.

Mol. Microbiol. 15, 267-275, 1995

A:title: Gonococcal rfaI mutants express Rd(2) chemotype LPS and do not enter epithel

A:Reference number: S61299; MUID:95264913

A:Accession: S61299

A&gt;Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-336 &lt;SCH&gt;

A:Cross-references: EMBL:Z27141; NID:9599919; PIDN:CAA85504.1; PID:9599920

A:Note: the sequence of residues 1-2 are not shown in this paper

C:Superfamily: ADP-heptose-LPS heptosyltransferase II

C:Keywords: glycosyltransferase

## alignment\_scores:

Quality:	50.00	Length:	29
Ratio:	2.632 <td>Gaps: <td>2</td> </td>	Gaps: <td>2</td>	2
Percent Similarity:	65.517 <td>Percent Identity:</td> <td>41.379</td>	Percent Identity:	41.379

## alignment\_block:

US-09-684-016-48411/rev x S61299 ..

Align seg 1/1 to: S61299 from: 1 to: 336

```

      113 AATAAATTGAAGATAGTATCACCTTACATGTGAACCTTACCAGTCAGTG 64
      ::|||
      293 AsparGalaLysIle...ValSerLeuHisLeuGlu...CysSerProCys 307
      |||
      63 CAGCCCTGCGCATTCGCGCATTCAGGAGGTTGT 27
      |||
      307 sPheLysArgGluCysProLeuGlyHisThrAspCys 319

```

seq\_name: p1r2:B81869

## seq\_documentation\_block:

lipopolysaccharide heptosyltransferase (EC 2.4.99.-) II KNA1727 [similarity] - *Neiss*C:Species: *Neisseria meningitidis*

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: B81869

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc



: Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
A:Reference number: AB1775; MUID:20222556  
A:Accession: B81869  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-336 <PAR>  
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84955.1; PID:g738036  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: rfaF; NMAL727  
C:Superfamily: ADP-heptose--LPS heptosyltransferase II  
C:Keywords: glycosyltransferase

alignment\_scores:  
Quality: 50.00 Length: 29  
Ratio: 2.632 Gaps: 2  
Percent Similarity: 65.517 Percent Identity: 41.379

alignment\_block:  
US-09-684-016-48411/rev x B81869 ..

Align seg 1/1 to: B81869 from: 1 to: 336

```
113 AATTAATTGAAGATAGTATCACCTTACATGTGAAGACTTACACAGTCAGTG 64
      ::::: ||||| :::::|||||::: || ||
293 AsparGAlaIle...ValSerLeuHisLeuGlu...CysSerProCy 307
      ||| :::||||| |||
63 CAGCCCTCGCATGCCGATTAAGTCATCAGGCTGT 27
      ||| :::||||| |||
307 spheIysArgGluCysProLeuGlyHisThrAspCys 319
```







```

seq_name: SwissProt_39:MOBA_NEIMA
seq_documentation_block:
ID MOBA_NEIMA STANDARD: PRT; 192 AA.
AC Q9JUA5:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE PROBABLE MOLYBDOPTERIN-GUANINE DINUCLEOTIDE BIOSYNTHESIS PROTEIN A.
GN MOBA OR NMA1417.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=63699;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;
MEDLINE=20222556; PubMed=10761919;
RA Parthill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
Whitehead S., Spratt B.G., Barrell B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491."
RT Nature 404:502-506(2000).
RL -1- FUNCTION: LINKS A GUANOSINE 5'-PHOSPHATE TO MOLYBDOPTERIN (MPT)
FORMING MOLYBDOPTERIN GUANINE DINUCLEOTIDE (MGD) (BY SIMILARITY).
CC -1- PATHWAY: MOLYBDENUM COPFACTOR BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MOBA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL162755; CAB84657.1;
KW Molybdenum cofactor biosynthesis; GTP-binding; Complete proteome.
SQ SEQUENCE 192 AA; 21803 MW; 005C495C6C64D290 CRC64;

alignment_scores:
Quality: 59.00 Length: 15
Ratio: 4.538 Gaps: 0
Percent Similarity: 86.667 Percent Identity: 66.667

alignment_block:
US-09-684-016-48411 x MOBA_NEIMA ..
Align seg 1/1 to: MOBA_NEIMA from: 1 to: 192
14 TGTGAGATGCTTACAACTGATGACCTAATGGCGCAATGCGAG 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 CYSAspMetProTyrLeuProAspAspLeuValAlaArgPheGlu 114

seq_name: SwissProt_39:FTSZ_BARBA
seq_documentation_block:
ID FTSZ_BARBA STANDARD: PRT; 592 AA.
AC O3J314:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CELL DIVISION PROTEIN FTSZ (75 KDA ANTIGEN).
GN FTSZ.
OS Bartonella bacilliformis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bartonellaceae; Bartonella.

```

```

OX NCBI_TaxID=774;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-KC584;
MEDLINE=97369823; PubMed=9226264;
RA Padmalayam I., Anderson B., Kiron M., Kelly T., Baumstark B.;
RT "The 75-kilodalton antigen of Bartonella bacilliformis is a
structural homolog of the cell division protein ftsZ."
RL J. Bacteriol. 179:4545-4552(1997).
RN [2]
RP REVISIONS.
RA Kelly T.;
RL Submitted (APR-1998) to the EMBL/Genbank/DDBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL TO THE CELL-DIVISION PROCESS.
ITS SEEMS TO ASSEMBLE INTO A DYNAMIC RING ON THE INNER SURFACE OF
THE CYTOPLASMIC MEMBRANE AT THE PLACE WHERE DIVISION WILL OCCUR,
AND THE FORMATION OF THE RING IS THE SIGNAL FOR SEPTATION TO
BEGIN. BINDS TO AND HYDROLYZES GTP (BY SIMILARITY).
CC -1- SUBUNIT: AGGREGATE TO FORM A RING-LIKE STRUCTURE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; ASSEMBLE AT THE INNER SURFACE
OF THE CYTOPLASMIC MEMBRANE (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF007266; AAC15082.1;
DR HSP: Q57816; FTSZ.
DR InterPro: IPR000158; FtsZ.
DR InterPro: IPR003008; Tubulin_FtsZ.
DR Pfam: PF00091; tubulin; 1.
DR PRINTS: PR00423; CELDIVISFTSZ.
DR PROSITE: PS01134; FTSZ_1;
DR PROSITE: PS01135; FTSZ_2; 1.
KW Cell division; Septation; GTP-binding.
FT NP_BIND 107 115 GTP (POTENTIAL).
SQ SEQUENCE 592 AA; 63503 MW; C32007DADCD2D75B CRC64;

alignment_scores:
Quality: 56.50 Length: 40
Ratio: 2.173 Gaps: 2
Percent Similarity: 65.000 Percent Identity: 32.500

alignment_block:
US-09-684-016-48411/rev x FTSZ_BARBA ..
Align seg 1/1 to: FTSZ_BARBA from: 1 to: 592
119 GTGTGATAATAATGAGATAGATACACTTA.....CAGTGAAC 76
:::|||||:|||||:|||||:|||||:|||||:|||||:
505 LeuTrpIleArgLeuIleGlnSerLeuThrTyrArgIleIleGluP 521
75 TACAGTCAAGTGCAGCCCTGCATGGCCATTAGGTACAGGATTGTA 26
:::::|||||:|||||:|||||:|||||:|||||:
521 OGAlaIArgLeuGluProAlaValIysProLeuGlnAsnGlu..... 535
25 AGGCATCTCACATTACAA 6
|||||:|||||:|||||:|||||:|||||:
536 ..GIuserHisIleTyrAsn 541

seq_name: SwissProt_39:GCKR_XENLA
seq_documentation_block:
ID GCKR_XENLA STANDARD: PRT; 619 AA.
AC Q91754:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)

```



DE GLUCOKINASE REGULATORY PROTEIN (GLUCOKINASE REGULATOR).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NX NCBI\_TaxID=8335;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=95010134; PubMed=7925465;  
 RA Veiga-Da-Cunha M., Delheux M., Watelet N., van Schaftingen E.;  
 RT Cloning and expression of a Xenopus liver cDNA encoding a fructose-  
 RT phosphate-insensitive regulatory protein of glucokinase.";  
 RL Eur. J. Biochem. 225:43-51(1994).  
 CC -1- FUNCTION: INHIBITS GLUCOKINASE BY FORMING AN INACTIVE COMPLEX WITH  
 CC THIS ENZYME.  
 CC -1- SIMILARITY: BELONGS TO THE SIS FAMILY. GCKR SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X80901; CAA56863.1; -  
 DR InterPro: IPR001741; GCKR.  
 DR InterPro: IPR001347; SIS.  
 DR Pfam: PF01380; SIS; 1.  
 DR PRODOM: PD025295; GCKR; 1.  
 DR PROSITE: PS01272; GCKR; 1.  
 SQ SEQUENCE 619 AA; 68738 MW; 41B72C1981D1BA52 CRC64;

alignment\_scores:  
 Quality: 56.50 Length: 28  
 Ratio: 2.568 Gaps: 1  
 Percent Similarity: 78.571 Percent Identity: 39.286

alignment\_block:

US-09-684-016-48411 x GCKR\_XENLA ..

Align seg 1/1 to: GCKR\_XENLA from: 1 to: 619

```

3 GGTATTGTAAT...GTGAGTCCTTACCACTGATGACCTAATGCGG 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 G|Y|L|E|E|T|S|P|A|L|A|V|A|I|G|U|C|Y|S|V|A|L|P|R|O|T|H|Y|G|I|N|A|L|A|S|P|R|A|R 357
50 CAATGCGAGGCGTGCCTGACTGCTGAGTGAAGTTTCAC 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||
357 G|A|S|P|V|A|L|G|I|P|H|E|I|E|H|G|I|G|Y|T|Y|T|H|S 368

```

seq\_name: SwissProt\_39:OPLA\_RAT

seq\_documentation\_block:

ID OPLA\_RAT STANDARD; PRT; 1288 AA.  
 AC P97608;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 5-OXOPROLINASE (EC 3.5.2.9) (5-OXO-L-PROLINASE) (PYROGLUTAMASE)  
 GN (5-OPASP).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NX NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;  
 RX MEDLINE=97113037; PubMed=8943290;  
 RA Ye G.-J., Breslow E., Meister A.;

RT "The amino acid sequence of rat kidney 5-oxo-L-prolinase determined  
 RT by cDNA cloning."  
 RL J. Biol. Chem. 271:32293-32300(1996).  
 RN (2)  
 RP ERRATUM.  
 RA Ye G.-J., Breslow E., Meister A.;  
 RL J. Biol. Chem. 272:4646-4646(1997).  
 CC -1- FUNCTION: CATALYZES THE CLEAVAGE OF 5-OXO-L-PROLINE TO FORM L-  
 CC GLUTAMATE COUPLED TO THE HYDROLYSIS OF ATP TO ADP AND INORGANIC  
 CC PHOSPHATE.  
 CC -1- CATALYTIC ACTIVITY: ATP + 5-OXO-L-PROLINE + 2 H(2)O = ADP +  
 CC PHOSPHATE + L-GLUTAMATE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- TISSUE SPECIFICITY: WELL EXPRESSED IN TESTIS, KIDNEY AND LIVER.  
 CC -1- SIMILARITY: BELONGS TO THE OXOPROLINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U70825; AAC52955.1; -  
 DR InterPro: IPR002821; Hydanolase\_A.  
 DR InterPro: IPR003692; Hydanolase\_B.  
 DR Pfam: PF01968; Hydanolase\_A; 1.  
 DR Pfam: PF02538; Hydanolase\_B; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 1288 AA; 137746 MW; 4530D0781E10C7AD CRC64;

alignment\_scores:  
 Quality: 56.00 Length: 30  
 Ratio: 2.800 Gaps: 1  
 Percent Similarity: 66.667 Percent Identity: 40.000

alignment\_block:

US-09-684-016-48411/rev x OPLA\_RAT ..

Align seg 1/1 to: OPLA\_RAT from: 1 to: 1288

```

86 CATGTGAACCTTACCAGTGCAGCCCT.....CGCATGGCCGCAAT 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||
215 H|I|S|V|A|L|S|E|R|U|S|E|R|S|E|R|G|I|V|A|L|M|E|R|O|M|E|T|Y|A|L|P|R|O|A|R 231
42 AGGTCAATCAGGGTGTGAAGCATCTCAATTTACAAATACC 3
|||||:|||||:|||||:|||||:|||||:|||||:|||||
231 G|G|Y|H|I|S|T|H|A|L|C|Y|S|A|L|A|S|P|A|L|A|T|Y|L|E|U|T|H|P|R|O|T|H|R 244

```

seq\_name: SwissProt\_39:H1P1\_HUMAN

seq\_documentation\_block:

ID H1P1\_HUMAN STANDARD; PRT; 995 AA.  
 AC 000291; O00328;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HUNTINGTIN INTERACTING PROTEIN 1 (HIP-1) (FRAGMENT).  
 GN H1P1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Hug A.H.M.M., Nichol K., Osborne L., Scherer S.W., Squitieri F.,  
 RA Hayden M.R.;  
 RT "Genomic organization of the human HIP1 gene and its exclusion as a  
 RT candidate gene in a family diagnosed with Huntington disease without  
 RT CAG expansion.";  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.







KA Grodeck E.O., Davis N.W., Linn A., Dimalanta E.T., Potamousis K.  
BA Apodaca J., Anantharaman T, S. Ijin T., Yen G., Schwartz D.C

0C Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
0C Escherichia



OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=93347969; PubMed=8346018;  
 RA Plunkett G., III, Burland V.D., Daniels D.L., Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. III. DNA sequence of the  
 RT region from 87.2 to 89.2 minutes.";  
 RL Nucleic Acids Res. 21:3391-3398(1993).  
 RN [2]  
 RP SEQUENCE OF 1-5, AND CHARACTERIZATION.  
 RX MEDLINE=94291668; PubMed=8020507;  
 RA Palmer T., Vasishta A., Whitty P.W., Boxer D.H.;  
 RT "Isolation of protein FA, a product of the mob locus required for  
 RT molybdenum cofactor biosynthesis in Escherichia coli.";  
 RL Eur. J. Biochem. 222:687-692(1994).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96004466; PubMed=7551035;  
 RA Llobi-Nivol C., Palmer T., Whitty P.W., McNaughton E., Boxer D.H.;  
 RT "The mob locus of Escherichia coli K12 required for molybdenum  
 RT cofactor biosynthesis is expressed at very low levels.";  
 RL Microbiology 141:1663-1671(1995).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).  
 RX MEDLINE=20568278; PubMed=10978347;  
 RA Lake M.W., Temple C.A., Rajagopalan K.V., Schindelin H.;  
 RT "The crystal structure of the Escherichia coli MobA protein provides  
 RT insight into molybdopterine guanine dinucleotide biosynthesis.";  
 RL J. Biol. Chem. 275:40211-40217(2000).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.35 ANGSTROMS).  
 RX PubMed=11080634;  
 RA Stevenson C.E., Sargent F., Buchanan G., Palmer T., Lawson D.M.;  
 RT "Crystal structure of the molybdenum cofactor biosynthesis protein  
 RT MobA from Escherichia coli at near-atomic resolution.";  
 RL Structure 8:1115-1125(2000).  
 CC -1- FUNCTION: LINKS A GUANOSINE 5'-PHOSPHATE TO MOLYBDOTERIN (MP)  
 CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.  
 CC -1- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE MOBA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L19201; AAB02992.1; -  
 DR EMBL: AE000461; AAC76855.1; -  
 DR PIR: S40803; S40803.  
 DR PDB: 1ESK; 11-NOV-00.  
 DR PDB: 1FR9; 10-JAN-01.  
 DR PDB: 1FRW; 10-JAN-01.  
 DR Ecogene: EG11829; mobA.  
 KW Molybdenum cofactor biosynthesis; GTP-binding; Magnesium;  
 KW 3D-structure; Complete proteome.  
 SO SEQUENCE 194 AA; 21643 MW; B/9B32DD7348DD48 CRC64;

alignment\_scores:  
 Quality: 52.00 Length: 16  
 Ratio: 4.727 Gaps: 0  
 Percent Similarity: 68.750 Percent Identity: 50.000

alignment\_block:  
 US-09-684-016-48411 x MOBA\_ECOLI

Align seg 1/1 to: MOBA\_ECOLI from: 1 to: 194  
 5 TATGTAAATGTCAGATGCTTACACCGTCGATGACCTAATGCGCAA 52  
 :||| ||||| ||| ||||| |||||  
 97 PhcysProcsAspThrProTyrIleProProAspLeuAlaIatay 112  
 seq\_name: SwissProt\_39:SUS2\_PEA  
 seq\_documentation\_block:  
 ID SUS2\_PEA STANDARD; PRT; 809 AA.  
 AC 024301;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SUCROSE SYNTHASE 2 (EC 2.4.1.13) (SUCROSE-UDP GLUCOSYLTRANSFERASE 2).  
 GN SUS2.  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
 OX NCBI\_Taxid=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. FRISSON; TISSUE=Seed coat;  
 RA Buchner P., Porel M., Rochat C.;  
 RT "Cloning and characterization of a cDNA encoding a second sucrose  
 RT synthase gene in pea (Pisum sativum L.).";  
 RL (in) Plant Gene Register PGR98-105.  
 CC -1- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND  
 CC FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.  
 CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE = UDP + SUCROSE.  
 CC -1- SIMILARITY: TO OTHER PLANTS SUCROSE SYNTHASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ001071; CA04512.1; -  
 DR InterPro: IPR001286; Glycos\_transf\_1.  
 DR InterPro: IPR000368; Sucrose\_synth.  
 DR Pfam: PF00534; Glycos\_transf\_1; 1.  
 DR Pfam: PF00862; Sucrose\_synth; 1.  
 KW Transferase; Glycosyltransferase; Multigene family.  
 SO SEQUENCE 809 AA; 92208 MW; 6C5BBF708C37DE75 CRC64;

alignment\_scores:  
 Quality: 52.00 Length: 57  
 Ratio: 2.364 Gaps: 2  
 Percent Similarity: 38.596 Percent Identity: 26.316

alignment\_block:

US-09-684-016-48411/rev x SUS2\_PEA

Align seg 1/1 to: SUS2\_PEA from: 1 to: 809

105 GAAGATAGATGATACCTTACATGTGAA..... 79  
 ||||| ||||| ||||| ||||| |||||  
 457 GluAspLysTyrHisPheSerCysGlnPheThrAlaAspLeuIleAla 473  
 78 ..... ACTTACGATGAGTG..... 64  
 ||||| ||||| ||||| ||||| |||||  
 473 tAsnAsnAlaAspPheIleIleThrSerThrTyrGlnGluIleAlaGlyT 490  
 63 ..... CAGCCCTGCGATGCGCCATAGGTCATCA 34  
 ||||| ||||| ||||| ||||| |||||  
 490 hLysAsnThrIleGlyGlnTyrGluSerHisThrAlaPheThrLeuPro 506  
 33 GCGTTGTAAGGATCTCACAT 13







```

seq_documentation_block:
ID  HEX9_ADECC  STANDARD:  PRT:  103 AA.
AC  Q65944;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  HEXON-ASSOCIATED PROTEIN (PROTEIN IX).
GN  PIX.
OS  Canine adenovirus type 1 (strain CLJ), and
OS  Canine adenovirus type 1 (strain RI261).
OC  Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX  NCBI_TaxID=69150, 69151;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CLJ;
RA  Campbell J.B., Zhao Y.;
RL  Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=RI261;
RA  MEDLINE=97275900; PubMed=9129661;
RA  Morrison M.D., Onions D.E., Nicolson L.;
RT  *Complete DNA sequence of canine adenovirus type 1.;
RL  J. Gen. Virol. 78:873-878(1997).
CC  -!- FUNCTION: THIS PROTEIN IS A STRUCTURAL COMPONENT OF THE VIRION. IT
CC  MAY HAVE AN ADDITIONAL ROLE DURING ADENOVIRUS MULTIPLICATION.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: U55001; AAB05432.1; -
DR  EMBL: Y07760; CAA69055.1; -
DR  Hexon-associated protein.
SQ  SEQUENCE 103 AA; 11303 MW; 4C14502AD6F5A8C1 CRC64;

alignment_scores:
Quality: 51.00 Length: 24
Ratio: 3.188 Gaps: 0
Percent Similarity: 66.667 Percent Identity: 50.000

alignment_block:
US-09-684-016-48411 x HEX9_ADECC
Align seg 1/1 to: HEX9_ADECC from: 1 to: 103

3 GGTATTGTAATGTGAGATGCTTACCACTGATGACCTAATGCGCAA 52
|||||
7 GYLIVeValasnthrcysPheleuthrThrArgIleproserIrrpAaGl 23
|||||
53 TCGAGGCGCTGCACGACTGCT 74
|||||
23 yAlaArGInaSnValThrGly 30
|||||

seq_name: Swissprot_39:DMSB_ECOLI
seq_documentation_block:
ID  DMSB_ECOLI  STANDARD:  PRT:  204 AA.
AC  P18776; P77745;
DT  01-NOV-1990 (Rel. 16, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  ANAEROBIC DIMETHYL SULFOXIDE REDUCTASE CHAIN B (EC 1.-.-.-) (DMSO
DE  REDUCTASE IRON-SULFUR SUBUNIT).
GN  DMSB OR B0895.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.

```

```

OX  NCBI_TaxID=562;
RN  [1]
RP  SEQUENCE FROM N.A., AND SEQUENCE OF 1-9.
RC  STRAIN=K12 / C600;
RC  MEDLINE=89096500; PubMed=3062312;
RA  Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H.;
RT  *Nucleotide sequence of the dmsABC operon encoding the anaerobic
RT  dimethylsulphoxide reductase of Escherichia coli.;
RL  Mol. Microbiol. 2:785-795(1988).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MG1655;
RC  MEDLINE=97426617; PubMed=9278503;
RA  Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Mau B., Shao Y.;
RT  *The complete genome sequence of Escherichia coli K-12.;
RL  Science 277:1453-1474(1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12;
RC  MEDLINE=97061202; PubMed=8905232;
RA  Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA  Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA  Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA  Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA  Sasaki G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA  Yano M., Horuchi T.;
RT  *A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT  corresponding to the 12.7-28.0 min region on the linkage map.;
RL  DNA Res. 3:1137-155(1996).
RN  [4]
RP  EPR SPECTROSCOPY OF IRON-SULFUR CLUSTERS.
RC  MEDLINE=91070067; PubMed=2174699;
RA  Cammack R., Weiner J.H.;
RT  *Electron paramagnetic resonance spectroscopic characterization of
RT  dimethyl sulfoxide reductase of Escherichia coli.;
RL  Biochemistry 29:8410-8416(1990).
RN  [5]
RP  MUTAGENESIS.
RC  MEDLINE=91355180; PubMed=1653010;
RA  Rothery R.A., Weiner J.H.;
RT  *Alteration of the iron-sulfur cluster composition of Escherichia
RT  coli dimethyl sulfoxide reductase by site-directed mutagenesis.;
RL  Biochemistry 30:8296-8305(1991).
CC  -!- FUNCTION: TERMINAL REDUCTASE DURING ANAEROBIC GROWTH ON
CC  VARIOUS SULFOXIDE AND N-OXIDE COMPOUNDS.
CC  -!- CORFACTOR: IRON-SULFUR.
CC  -!- SUBUNIT: THE COMPLEX CONSIST OF THREE SUBUNITS: DMSA, THE
CC  REDUCTASE; DMSB, AN ELECTRON TRANSFER PROTEIN, AND DMSB, A
CC  MEMBRANE ANCHOR PROTEIN.
CC  -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC  'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: J03412; AAA83844.1; -
DR  EMBL: AE000191; AAC73981.1; -
DR  EMBL: D90727; BA35627.1; -
DR  FTR: S03786; S03786.
DR  HSP: P00195; ICLF.
DR  EcoGene: EG10233; dmsB.
DR  InterPro: IPR001450; 4FE4S_ferredoxin.
DR  Pfam: PF00037; fer4_1.
DR  PROSITE: PS00198; 4FE4S_FERREDOXIN; 1.
DR  Oxidoreductase; Electron transport; 4Fe-4S; Iron-sulfur;
KW

```



KW Complete proteome. 0  
 FT INIT MER 0 0  
 FT METAL 13 13 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY)  
 FT METAL 16 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY)  
 FT METAL 19 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY)  
 FT METAL 23 23 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY)  
 FT METAL 66 66 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY)  
 FT METAL 69 69 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY)  
 FT METAL 74 74 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY)  
 FT METAL 78 78 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY)  
 FT METAL 98 98 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY)  
 FT METAL 101 101 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY)  
 FT METAL 104 104 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY)  
 FT METAL 108 108 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)  
 FT METAL 125 125 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)  
 FT METAL 128 128 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)  
 FT METAL 140 140 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)  
 FT METAL 144 144 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)  
 FT MUTAGEN 101 101 C->F/S/W/Y: LOSS OF ELECTRON TRANSFER FROM MENAQUINOL TO DMSO.  
 FT CONFLICT 169 169 P -> PRA (IN REF. 1).  
 SQ SEQUENCE 204 AA; 22738 MW; 90425D307DEF2130 CRC64;

alignment\_scores:  
 Quality: 51.00 Length: 29  
 Ratio: 3.400 Gaps: 1  
 Percent Similarity: 51.724 Percent Identity: 34.483

alignment\_block:

US-09-684-016-48411 x DMSB\_ECOLI ..

Align seg 1/1 to: DMSB\_ECOLI from: 1 to: 204

2 AGTATGTGTAATGTCAGATCCCTTACACCCCTGATGACCTA..... 43  
 |||||  
 102 ATGTCTCTGSHmetalaCysProtYGLYAlaProGlnInTyRasnGluTh 118  
 44 .....ATGGCGCATGCGAGCGCTGCACATGAC 70  
 118 rlysglyNhisMetThrlyscysasprglycystyrasp 130

seq\_name: SwissProt\_39: PUR8\_CHICK

seq\_documentation\_block:

ID PUR8\_CHICK STANDARD; PRT; 459 AA.

AC P21263;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ADENYLOSUCCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCINASE) (ASL) (ASASE).  
 GN ADSL.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=90264380; PubMed=2111814;  
 RA Aimi J., Badylak J., Williams J., Chen Z., Zalkin H., Dixon J.E.;  
 RT "Cloning of a cDNA encoding adenylosuccinate lyase by functional  
 complementation in *Escherichia coli*."  
 RL J. Biol. Chem. 265:9011-9014(1990).  
 CC -1- CATALYTIC ACTIVITY: 1-(5'-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXYAMIDE)  
 CC -5-AMINOIMIDAZOLE = FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-  
 CC IMIDAZOLECARBOXYAMIDE (ALSO CATALYZES: N6-(1,2-DICARBOXYETHYL)AMP =  
 CC FUMARATE + AMP)  
 CC -1- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE LYASE I FAMILY. ADENYLOSUCCINATE LYASE  
 CC SUBFAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M37901; AAA48574.1; -  
 DR PIR: A35291; A35291.  
 DR InterPro: IPR000362; Fumarate\_lyase.  
 DR Pfam: PF00206; Lyase\_1; 1.  
 DR PRINTS: PR00149; FUMARATELYASE.  
 DR PROSITE: PS00163; FUMARATE\_LYASES; 1.  
 KW Purine biosynthesis; Lyase.  
 FT ACT\_SITE 61 61 ACID (BY SIMILARITY).  
 FT ACT\_SITE 134 134 BASE (BY SIMILARITY).  
 SQ SEQUENCE 459 AA; 51786 MW; E5EC141F69BCBAC5 CRC64;

alignment\_scores:  
 Quality: 51.00 Length: 33  
 Ratio: 2.550 Gaps: 0  
 Percent Similarity: 60.606 Percent Identity: 27.273

alignment\_block:

US-09-684-016-48411/rev x PUR8\_CHICK ..

Align seg 1/1 to: PUR8\_CHICK from: 1 to: 459

110 AATTTGAAGATGAATGATACCTTACATGCTGAACCTTACAGTCAGTCAG 61  
 |||||  
 58 lysleuargNhisAsprValMetAlaHisValHisThrAlaHisCysCys 74  
 60 CCTCGCATGTCGCGCATTAAGTATGATCAAGGTTGTGAAGCATGTACATT 12  
 |||||  
 74 sprlyAlaAlaAlaIleIleHisLeuGlyAlaThrSerCysTyrVal 90

seq\_name: SwissProt\_39:E315\_ADE05

seq\_documentation\_block:

ID E315\_ADE05 STANDARD; PRT; 132 AA.

AC P06498;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE EARLY E3B 14.6 KDA PROTEIN PRECURSOR.  
 OS Human adenovirus type 5.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OC NCBI\_TaxID=28285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85092388; PubMed=2981456;  
 RA Cladaras C., Wold W.S.M.;  
 RT "DNA sequence of the early E3 transcription unit of adenovirus 5."  
 RL Virology 140:28-43(1985).  
 RN [2]  
 RP IDENTIFICATION OF PROTEIN.  
 RP MEDLINE=90117214; PubMed=2309441;  
 RA Tollefson A.E., Krajsi P., Pursley M.H., Gooding L.R., Wold W.S.M.;  
 RT "A 14,500 MW protein is coded by region E3 of group C human  
 adenoviruses."  
 RL Virology 175:19-29(1990).  
 RN [3]  
 RP SIGNAL SEQUENCE CLEAVAGE SITE.  
 RP MEDLINE=92148933; PubMed=1531370;  
 RX Krajsi P., Tollefson A.E., Anderson C.W., Wold W.S.M.;  
 RA "The adenovirus E3 14.5-kilodalton protein, which is required for  
 RT down-regulation of the epidermal growth factor receptor and  
 RT prevention of tumor necrosis factor cytotoxicity, is an integral  
 RT membrane protein oriented with its C terminus in the cytoplasm."  
 RL J. Virol. 66:1665-1673(1992).  
 RN [4]



```

RP PHOSPHORYLATION.
RX MEDLINE=92188523; PubMed=1532104;
RA Krajesi P., Wold W.S.M.;
RT "The adenovirus E3-14.5K protein which is required for prevention of
RT TNF cytotoxicity and for down-regulation of the EGF receptor contains
RT phosphoserine."
RL Virology 187:492-498(1992).
RN [5]
RP O-GLYCOSYLATION.
RX MEDLINE=92263760; PubMed=1533979;
RA Krajesi P., Tollefson A.E., Wold W.S.M.;
RT "The E3-14.5K integral membrane protein of adenovirus that is
RT required for down-regulation of the EGF receptor and for prevention
RT of TNF cytotoxicity is O-glycosylated but not N-glycosylated."
RL Virology 188:570-579(1992).
RN [6]
RP COMPLETE GENOME.
RX MEDLINE=92087470; PubMed=1727603;
RA Chroboczek J., Biebler F., Jacrot B.;
RT "The sequence of the genome of adenovirus type 5 and its comparison
RT with the genome of adenovirus type 2."
RL Virology 186:280-285(1992).
CC -1- FUNCTION: DOWN-REGULATES THE EGF RECEPTOR AND PREVENTS CYTOLYSIS
CC BY TNF.
CC -1- PTM: PHOSPHORYLATED ON SERINE; O-GLYCOSYLATED, BUT NOT
CC N-GLYCOSYLATED.
CC -1- SIMILARITY: TO EQUIVALENT PROTEIN IN OTHER ADENOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73260; -; NOT_ANNOTATED_CDS.
DR EMBL; X03002; CA26786.1; -.
KW Early protein; Transmembrane; Glycoprotein; Phosphorylation; Signal.
FT SIGNAL 1 18
FT CHAIN 19 132 EARLY E3B 14.6 KDA PROTEIN.
FT TRANSMEM 51 75 POTENTIAL.
FT SEQUENCE 132 AA; 14750 MW; 524690C4AD9B9A74 CRC64;

alignment_scores:
Quality: 50.50 Length: 37
Ratio: 2.196 Gaps: 1
Percent Similarity: 62.162 Percent Identity: 29.730

alignment_block:
US-09-684-016-48411/rev x E315_ADE05 ..
Align seg 1/1 to: E315_ADE05 from: 1 to: 132

107 TTGAAGATAGATATACAC.....TTACATGTGAACACTTACACACTCA 67
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetLysPheThrValThrPheLeuLeuIleIleCysThrLeuSerAlaIap 17
66 GAGCAGCCCTGGCATTTGGCGCATTCAGCGGTGTGAAGGATCTTC 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 eCysSerProThrSerLysPProGlnArgHisIleSerCysArgPheThr 34
16 ACATTTCACAT 6
:|||||:|||||:|||||:|||||:|||||:|||||:
34 gLleTrrPasn 37

seq_name: SwissProt_39:SPC2_CHICK
seq_documentation_block:
ID SPC2_CHICK STANDARD; PRT; 180 AA.
AC P28687;
DT 01-DEC-1992 (Rel. 24, Created)

```

```

DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MICROSOMAL SIGNAL PEPTIDASE 23 KDA SUBUNIT (EC 3.4.-.-) (SPC22/23)
DE (GP23).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Oviduct;
RX MEDLINE=92189580; PubMed=1546959;
RA Newsome A.L., McLean J.W., Lively M.O.;
RT "Molecular cloning of a cDNA encoding the glycoprotein of hen oviduct
RT microsome signal peptidase."
RL Biochem. J. 282:447-452(1992).
CC -1- FUNCTION: MICROSOMAL SIGNAL PEPTIDASE IS A MEMBRANE-BOUND
CC ENDOPEPTIDASE THAT REMOVES SIGNAL PEPTIDES FROM NASCENT
CC PROTEINS AS THEY ARE TRANSLOCATED INTO THE LUMEN OF THE
CC ENDOPLASMIC RETICULUM.
CC -1- SUBUNIT: HEN OVIDUCT SIGNAL PEPTIDASE ACTIVITY IS ASSOCIATED
CC WITH AN APPARENT COMPLEX OF ONLY TWO PROTEINS, WITH ESTIMATED
CC MOLECULAR MASSES OF 19 AND 23 KDA.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MICROSOMAL.
CC -1- SIMILARITY: TO OTHER VERTEBRATES SPC22/23 AND TO YEAST SPC3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60795; CAA43208.1; -.
DR PIR; S15551; S15551.
DR PIR; S22412; S22412.
KW Glycoprotein; Hydrolase; Microsome; Endoplasmic reticulum;
KW Transmembrane; Protease; Signal-anchor.
FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 11 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 33 180 (POTENTIAL).
FT DOMAIN 144 164 LUMENAL (POTENTIAL).
FT CARBOHYD 141 141 HYDROPHOBIC.
FT SEQUENCE 180 AA; 20230 MW; A5572439EB8210AA CRC64;

alignment_scores:
Quality: 50.00 Length: 28
Ratio: 2.500 Gaps: 0
Percent Similarity: 71.429 Percent Identity: 39.266

alignment_block:
US-09-684-016-48411 x SPC2_CHICK ..
Align seg 1/1 to: SPC2_CHICK from: 1 to: 180

38 GACCTATATGGCGAATGCGAGGCTGCACCTGATGTTACATCT 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 AsplleThrAlaSplleunSerIlePheAspTrrPasnValIyselnde 85
88 AAGGTGATACCTTATCTTCAATTATTCACACA 121
:|||||:|||||:|||||:|||||:|||||:|||||:
85 upheLeuTyrrLeuSerAlaGlnTyrrSerThrIys 96

seq_name: SwissProt_39:NGFR_HUMAN
seq_documentation_block:
ID NGFR_HUMAN STANDARD; PRT; 427 AA.
AC P08138;
DT 01-AUG-1988 (Rel. 08, Created)

```



DT 01-AUG-1988 (Rel. 08, last sequence update)  
 DT 20-AUG-2001 (Rel. 40, last annotation update)  
 DE LOW-AFFINITY NERVE GROWTH FACTOR RECEPTOR PRECURSOR (NGF RECEPTOR)  
 DE (GP80-LINGER) (P75 ICD).  
 GN NGFR OR TNFRSF16.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87051725; PubMed=3022937;  
 RA Johnson D., Lanthan A., Buck C.R., Sehgal A., Morgan C., Mercer E.,  
 RA Bothwell M., Chao M.;  
 RT "Expression and structure of the human NGF receptor.";  
 RL Cell 47:545-554(1986).  
 RN [2]  
 RP SEQUENCE OF 1-22 FROM N.A.  
 RX MEDLINE=89096903; PubMed=2850481;  
 RA Sehgal A., Patil N., Chao M.;  
 RT "A constitutive promoter directs expression of the nerve growth factor  
 RT receptor gene.";  
 RL Mol. Cell. Biol. 8:3160-3167(1988).  
 CC -1- FUNCTION: LOW AFFINITY RECEPTOR WHICH CAN BIND TO NGF, BDNF,  
 CC NT-3, AND NT-4.  
 CC -1- SUBUNIT: NGF RECEPTOR CAN FORM A HOMODIMER THROUGH DISULFIDE  
 CC BOND FORMATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- PTM: N- AND O-GLYCOSYLATED AND IS PHOSPHORYLATED ON SERINE.  
 CC -1- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.  
 CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL M14764; AAB59544.1; -  
 DR EMBL M21621; AAA36363.1; -  
 DR PIR A25218; GOHUN.  
 DR HSSP P07174; INGR.  
 DR MIM 162010; -  
 DR InterPro: IPR000488; Death.  
 DR InterPro: IPR001368; TNFR\_cf.  
 DR Pfam: PF00531; death; 1.  
 DR Pfam: PF00020; TNFR\_c6; 4.  
 DR SMART: SM00005; DEATH; 1.  
 DR SMART: SM00208; TNFR; 3.  
 DR PROSITE: PS00652; TNFR\_NGFR\_1; 3.  
 DR PROSITE: PS50050; TNFR\_NGFR\_2; 4.  
 DR PROSITE: PS50017; DEATH\_DOMAIN; 1.  
 KM Receptor; Neurogenesis; Transmembrane; Glycoprotein; Repeat;  
 KM Phosphorylation; Signal.  
 FT SIGNAL 1 28  
 FT CHAIN 29 427  
 FT DOMAIN 29 250  
 FT TRANSMEM 251 272  
 FT DOMAIN 273 427  
 FT DOMAIN 31 189  
 FT REPEAT 31 65  
 FT REPEAT 66 107  
 FT REPEAT 108 147  
 FT REPEAT 148 189  
 FT DOMAIN 344 421  
 FT DISULFID 32 43  
 FT DISULFID 44 57  
 FT DISULFID 47 64  
 FT DISULFID 67 83  
 FT DISULFID 86 99  
 CC BY SIMILARITY.

FT DISULFID 89 107 BY SIMILARITY.  
 FT DISULFID 109 122 BY SIMILARITY.  
 FT DISULFID 125 138 BY SIMILARITY.  
 FT DISULFID 128 146 BY SIMILARITY.  
 FT DISULFID 149 164 BY SIMILARITY.  
 FT DISULFID 167 180 BY SIMILARITY.  
 FT DISULFID 170 188 BY SIMILARITY.  
 FT DOMAIN 197 248 BY SIMILARITY.  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC... (POTENTIAL).  
 SQ SEQUENCE 427 AA; 45183 MW; B09FA143FB3D625B CRC64;  
 alignment\_scores: Length: 19  
 Quality: 50.00 Gaps: 0  
 Ratio: 4.167  
 Percent Similarity: 63.158 Percent Identity: 36.842  
 alignment\_block:  
 US-09-684-016-48411 x NGFR\_HUMAN ..  
 Align seg 1/1 to: NGFR\_HUMAN from: 1 to: 427  
 8 TGTAAATGTGAGAGCGCTTACACCCGTATGACCTAATGGCGCAATGCGA 57  
 |||:||||| ||| :|||:|||||  
 107 CysatgCysAlatYtYgltYtYtYtGlnaspGluThrThrGlyArgCysGcl 123  
 58 GGGCTGC 64  
 |||:|||||  
 123 uAlacys 125  
 seq\_name: SwissProt\_39:PHF1\_MOUSE  
 seq\_documentation\_block:  
 ID PHF1\_MOUSE STANDARD; PRT; 559 AA.  
 AC Q921B8; Q54808;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, last sequence update)  
 DT 20-AUG-2001 (Rel. 40, last annotation update)  
 DE PHD FINGER PROTEIN 1 (PHF1 PROTEIN) (T-COMPLEX TESTIS-EXPRESSED 3)  
 DE (POLYCOMBLIKE 1) (MECL1).  
 GN PHF1 OR TCTEX3 OR TCTEX-3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NIH SWISS;  
 RA Howard T.L., Ingermann A.R., Hollenberg S.M.;  
 RT "The basic helix-loop-helix protein Tth interacts with polycomb-group  
 RT proteins.";  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=99018220; PubMed=9799836;  
 RA Kawakami S., Mitsuunaga K., Kikuti Y.Y., Ando A., Inoko H.,  
 RA Yamamura K., Abe K.;  
 RT "Tctex3, related to Drosophila polycomblike, is expressed in male germ  
 RT cells and mapped to the mouse t-complex.";  
 RL Mamm. Genome 9:874-880(1998).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: TESTIS-SPECIFIC.  
 CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----



DR EMBL: U81490: AAD00518.1; -  
 DR EMBL: AB011550: BAA25074.1; -  
 DR MGI: 98647: Tctex3.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR002999; Tudor.  
 DR Pfam: PF00628; PHD; 2.  
 DR SMART: SM00249; PHD; 2.  
 DR SMART: SM00333; Tudor; 1.  
 DR Nuclear protein; Zinc-finger; Repeat.  
 FT ZN\_FING 89 142 PHD-type 1.  
 FT ZN\_FING 188 237 PHD-type 2.  
 FT CONFLICT 249 249 L -> F (IN REF. 1).  
 FT CONFLICT 344 344 G -> S (IN REF. 2).  
 FT CONFLICT 377 377 R -> S (IN REF. 2).  
 FT CONFLICT 555 555 G -> R (IN REF. 2).  
 FT CONFLICT 558 559 IF -> HLPDLSLLLPSPPTHHFHALDL (IN REF. 2).  
 SO SEQUENCE 559 AA; 61139 MM; 6D2E5F53D6164C2 CRC64;

## alignment\_scores:

Quality: 50.00 Length: 23  
 Ratio: 3.571 Gaps: 0  
 Percent Similarity: 60.870 Percent Identity: 39.130

## alignment\_block:

US-09-684-016-48411 x PHF1\_MOUSE ..

Align seg 1/1 to: PHF1\_MOUSE from: 1 to: 559

5 TATGTGAATGATGATGCTTACACCGTACGATGACCTATGGCGCATG 54  
 ||||| ||| ||| : : : ||| ||| |||||  
 188 Tyrcystyrcsglyglyproglyglutrpasnleuylswetleuincy 204  
 55 CGAGGCTGCACGACTGCTG 73  
 | : : : : ||| : : : |||  
 204 sargsercysleuglntrp 210

seq\_name: Swissprot\_39:PHF1\_HUMAN

## seq\_documentation\_block:

ID PHF1\_HUMAN STANDARD: PRT; 567 AA.  
 AC 043189; 060929;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PHD FINGER PROTEIN 1 (PHF1 PROTEIN).  
 GN PHF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Placenta;  
 RX MEDLINE=98207256; PubMed=9545646;  
 RA Coulson M., Robert S., Eyre H.J., Saint R.;  
 RT "The identification and localization of a human gene with sequence  
 RT similarity to Polycomblike of Drosophila melanogaster.";  
 RT Genomics 48:381-383(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RA Wang J.H., Du G.W., Zhou Y., Yuan J.G., Qiang B.O.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RA Mashreghi-Mohammadi M.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/PHF1 AND 2/PHF2 (SHOWN HERE);  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN HEART, SKELETAL MUSCLE, AND  
 CC PANCREAS, LOWER LEVELS IN BRAIN, PLACENTA, LONG, LIVER, AND

CC KIDNEY.  
 CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).

CC EMBL: AF029678: AAC52062.1; -  
 CC EMBL: AF052205: AAC13273.1; -  
 CC EMBL: AL021366: CAA16158.1; -  
 CC EMBL: AL021366: CAA16159.1; -  
 CC MIM: 602881; -  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR002999; Tudor.  
 DR Pfam: PF00628; PHD; 2.  
 DR SMART: SM00249; PHD; 2.  
 DR SMART: SM00333; Tudor; 1.  
 DR Nuclear protein; Zinc-finger; Repeat; Alternative splicing.  
 FT ZN\_FING 89 142 PHD-type 1.  
 FT ZN\_FING 188 237 PHD-type 2.  
 FT VARSPLIC 350 457  
 FT SPPSGCGGCGVSRPLGRRRPPPEPLRRRQKVEELGPP  
 FT SAVNRQPEQDERAHLOALQASVSPSPSPNOYOGSS  
 FT GYNERPTDRCILPSPDIRMFASFHS -> RAGPWGRLTS  
 FT PGEAPFAGARAPDEEAGEGSCGAGATLSSAQSARAPAGEA  
 FT GSSAEGTAAPSGCLPSTLLPAPQGLGTVDPTQGHMWN  
 FT TLVSPQSLKVPPT (IN ISOFORM 1).  
 FT VARSPLIC 458 567 MISSING (IN ISOFORM 1).  
 SO SEQUENCE 567 AA; 62077 MM; 3492A94ADP959579 CRC64;

## alignment\_scores:

Quality: 50.00 Length: 23  
 Ratio: 3.571 Gaps: 0  
 Percent Similarity: 60.870 Percent Identity: 39.130

## alignment\_block:

US-09-684-016-48411 x PHF1\_HUMAN ..

Align seg 1/1 to: PHF1\_HUMAN from: 1 to: 567

5 TATGTGAATGATGATGCTTACACCGTACGATGACCTATGGCGCATG 54  
 ||||| ||| ||| : : : ||| ||| |||||  
 188 Tyrcystyrcsglyglyproglyglutrpasnleuylswetleuincy 204  
 55 CGAGGCTGCACGACTGCTG 73  
 | : : : : ||| : : : |||  
 204 sargsercysleuglntrp 210

seq\_name: Swissprot\_39:BS4\_HUMAN

## seq\_documentation\_block:

ID BS4\_HUMAN STANDARD: PRT; 601 AA.  
 AC Q9Y5A7; 095422;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE BS4 PROTEIN (NY-REN-18 ANTIGEN).  
 GN NYREN18.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99438124; PubMed=10508479;  
 RA Scallan M.J., Jordan J.D., Williamson B., Stockert E., Bander N.H.,  
 RA Jorgensen V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,  
 RA Old L.J.;  
 RT "Antigens recognized by autologous antibody in patients with renal-



RT cell carcinoma.";  
 RL Int. J. Cancer 83:456-464(1999).  
 RN [2]  
 RP SEQUENCE OF 1-365 FROM N.A.  
 RA Rump A., Rosenthal A., Drescher B., Weber J., Schatevov R.,  
 RA Korenberg J.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 3 UBA DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF150509; AAD42865.1; -  
 DR EMBL: AF108083; AAC82474.1; -  
 DR InterPro: IPR000449; UBA.  
 DR Pfam: PF00627; UBA; 2.  
 DR SMART: SM00165; UBA; 1.  
 KW Repeat.  
 FT DOMAIN 374 414 UBA 1.  
 FT DOMAIN 424 456 UBA 2.  
 FT DOMAIN 475 515 UBA 3.  
 FT CONFLICT 75 75  
 FT D -> DNYRTGATIEVPLPRKDKIAKQVSDRLCCEN  
 FT CONFLICT 365 365 EVEKVEIEIRCKAIEGTGND (IN REF. 2).  
 FT R -> K (IN REF. 2).  
 SQ SEQUENCE 601 AA; 69272 MM; D5A6ECAE0B3A975B CRC64;

## alignment\_scores:

Quality: 50.00 Length: 46  
 Ratio: 2.174 Gaps: 2  
 Percent Similarity: 50.000 Percent Identity: 36.957

## alignment\_block:

US-09-684-016-48411/rev x BS4\_HUMAN ..

Align seg 1/1 to: BS4\_HUMAN from: 1 to: 601

```

113 AATAAATGAGATAGATACCTTACATGTGAACCTTACCAGTCAGTG 64
      ::::::::::: ||| ::::| ||| |||
368 SerSerLeuLysSerTyrIleLeuIleHisGlnYstrPhrIleCysCy 384
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
63 CAGC.....CCTGCATTTGGCGCATTAAGTCATCAGGGTGT. 27
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
384 sSerTrpGlyLeuLeuProArgLysHisArgLeuGlyLeuArgAlaCysA 401
      26 .....AAGCATCTCACATTTACAAT 6
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 spGlyAsnValAspHisAlaIleThrHisIleThrAsn 413

```

seq\_name: SwissProt\_39:SUSY\_BETVU

## seq\_documentation\_block:

ID SUSY\_BETVU STANDARD; PRT; 766 AA.  
 AC 042652;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SUCROSE SYNTHASE (EC 2.4.1.13) (SUCROSE-UDP GLUCOSYLTRANSFERASE)  
 DE (FRAGMENT).  
 GN SSI.  
 OS Beta vulgaris (Sugar beet).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.  
 OC NCBI\_TaxID=3555;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Tap root;  
 RC

RX MEDLINE=96270366; PubMed=8639746;  
 RA Hesse H., Willmitzer L.;  
 RA "Expression analysis of a sucrose synthase gene from sugar beet (Beta  
 RT vulgaris L.)";  
 RL plant Mol. Biol. 30:863-872(1996).  
 CC -1- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND  
 CC FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.  
 CC CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE = UDP + SUCROSE.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED MOST PREDOMINANTLY IN TAP ROOT.  
 CC -1- SIMILARITY: TO OTHER PLANTS SUCROSE SYNTHASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X81974; CAA57499.1; -  
 DR InterPro: IPR001296; Glycos\_transf\_1.  
 DR InterPro: IPR000368; Sucrose\_synth.  
 DR Pfam: PF00534; Glycos\_transf\_1; 1.  
 DR Pfam: PF00862; Sucrose\_synth; 1.  
 DR Transferase; Glycosyltransferase.  
 FT NON TER 1  
 FT 1  
 SQ SEQUENCE 766 AA; 87240 MM; 407803BCD8BC4A82 CRC64;

## alignment\_scores:

Quality: 50.00 Length: 57  
 Ratio: 2.273 Gaps: 2  
 Percent Similarity: 38.596 Percent Identity: 26.316

## alignment\_block:

US-09-684-016-48411/rev x SUSY\_BETVU ..

Align seg 1/1 to: SUSY\_BETVU from: 1 to: 766

```

105 GAGATAGATACCTTACATGTGA..... 79
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 GluAspLysTyrHisPheSerCysGlnPheSerAlaAspLeuMetAla 415
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78 .....ACTTACCAGTCAGG..... 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
415 tAsnHisAlaAspPheIleIleThrSerThrTyrGlnIleValIleAlaGlyT 432
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
63 .....CAGCCCTGCATTTGGCCCATTAAGTCATCA 34
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 hTrLysAsnThrValGlyGlnTyrGluSerHisLysAlaPheThrPhePro 448
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
33 GGGTTGTAAAGCATCTCACAT 13
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 GlyLeuTyrArgValValHis 455

```

seq\_name: SwissProt\_39:BAR3\_CHITE

## seq\_documentation\_block:

ID BAR3\_CHITE STANDARD; PRT; 1700 AA.  
 AC 003376;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE BALBIANT RING PROTEIN 3 PRECURSOR.  
 GN BR3.  
 OS Chironomus tentans (Midge).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
 OC Chironomidae; Chironomidae; Chironominae; Chironomus.  
 OC NCBI\_TaxID=7153;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Salivary gland;  
 RC



```

RX MEDLINE-90172404: PubMed-1689777;
RA Paulsson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
RL repetitive structure split by many introns.";
J. Mol. Biol. 211:331-349(1990).
CC -1- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
CC STRUCTURE. THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SALIVARY GLAND.
CC -1- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52263; CAA36506.1; .
DR PIR: S08167; S08167.
DR HSP: P18055; 2MRB.
DR InterPro: IPR000853; M11111ion_nemat.
DR PRINTS: PR00876; MTNEMATODE.
DR Repeat: Signal.
FT SIGNAL 1
FT CHAIN ? 1700 BALBIANI RING PROTEIN 3.
SQ SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;

alignment_scores:
  Quality: 50.00      Length: 22
  Ratio: 4.545        Gaps: 1
  Percent Similarity: 50.000  Percent Identity: 45.455

alignment_block:
US-09-684-016-48411 x BAR3_CHITE ..
Align seg 1/1 to: BAR3_CHITE from: 1 to: 1700

      8 TGTAAATGTGAGATGCGCTTACACACCTGATGACCTAATGGCGCATGCGA 57
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      1175 CysGlnCysGlnCysPro.....LysAspLysProLysLysGlnCysSpr 1189
      58 GGGCTGCACTGACTCG 73
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      1189 oGlyGlyGlnAspTTP 1194

seq_name: SwissProt_39:RBB2_HUMAN

seq_documentation_block:
ID RBB2_HUMAN STANDARD: PRT: 1722 AA.
AC P29375;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETINOBLASTOMA BINDING PROTEIN 2 (RBBP-2).
GN RBBP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94020841; PubMed-8414517;
RA Fattley A.R., Helin K., Dembski M.S., Dyson N., Harlow E.,
RA Vucocolo G.A., Hanobik M.G., Haskell K.M., Oliff A., Defeo-Jones D.,
RA Jones R.E.;
RT "Characterization of the retinoblastoma binding proteins RBP1 and
RBP2.";

```

```

RL Oncogene 8:3149-3156(1993).
RN [2]
RP SEQUENCE OF 1102-1569 FROM N.A.
RX MEDLINE-91312450; PubMed-1857421;
RA Defeo-Jones D., Huang P.S., Jones R.E., Haskell K.M., Vucocolo G.A.,
RA Hanobik M.G., Huber H.E., Oliff A.;
RT "Cloning of cDNAs for cellular proteins that bind to the
RT retinoblastoma gene product.";
RL Nature 352:251-254(1991).
CC -1- FUNCTION: INTERACTS WITH THE VIRAL PROTEIN-BINDING DOMAIN OF THE
CC RETINOBLASTOMA PROTEIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO THE C-TERMINAL DOMAIN OF THE XE169 PROTEIN AND
CC THE JUMONJI PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S66431; AAB28544.1; .
DR PIR: S16954; S16954.
DR MIM: 180202; .
DR InterPro: IPR001606; ARID.
DR InterPro: IPR003347; JmjC.
DR InterPro: IPR003349; JmjN.
DR InterPro: IPR001965; PHD.
DR Pfam: PF01388; ARID; 1.
DR Pfam: PF02373; JmjC; 1.
DR Pfam: PF02375; JmjN; 1.
DR Pfam: PF00628; PHD; 3.
DR SMART: SM00501; BRIGHT; 1.
DR SMART: SM00545; JmjN; 1.
DR SMART: SM00249; PHD; 3.
DR Trans-acting factor; Nuclear protein.
FT CONFLICT 1563 1566 MISSING (IN REF. 2).
FT SEQUENCE 1722 AA; 195815 MW; 8CF8A8A8E69A652 CRC64;

alignment_scores:
  Quality: 50.00      Length: 24
  Ratio: 3.333        Gaps: 1
  Percent Similarity: 62.500  Percent Identity: 37.500

alignment_block:
US-09-684-016-48411 x RBB2_HUMAN ..
Align seg 1/1 to: RBB2_HUMAN from: 1 to: 1722

      2 AGGATGTGAATGTGAGATGCGCTTACACACCTGATGACCTAATGGCGCA 51
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      1162 LysPheCysIleCysArgLys.....ThrAlaSerGlyPheMetLeuGI 1176
      52 ATGGAGAGCGCTGCACCTGACTCG 73
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      1176 nCysGluLeuCysLysAspTTP 1183

seq_name: SwissProt_39:COBJ_PSEDE

seq_documentation_block:
ID COBJ_PSEDE STANDARD: PRT: 254 AA.
AC P21640;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PRECORIN-3B C17-METHYLTRANSFERASE (EC 2.1.1.131) (PRECORIN-3
DE METHYLTRANSFERASE) (PRECORIN-3 METHYLASE).
GN COBJ.
OS Pseudomonas denitrificans.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

```



```

OC Pseudomonas.
OX NCBI_TaxID=43306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC510;
RA MEDLINE=91008976; PubMed=2211521;
RA Crouzet J., Cameron B., Cauchois L., Rigault S., Rouyez M.-C.,
RA Blanche F., Thibaut D., Debussche L.;
RT "Genetic and sequence analysis of an 8.7-kilobase Pseudomonas
RT denitrificans fragment carrying eight genes involved in
RT transformation of precorrin-2 to cobyrinic acid.";
RL J. Bacteriol. 172:5980-5990(1990).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=94042919; PubMed=8226690;
RA Debussche L., Thibaut D., Cameron B., Crouzet J., Blanche F.;
RT "Biosynthesis of the corrin macrocycle of coenzyme B12 in Pseudomonas
RT denitrificans.";
RL J. Bacteriol. 175:7430-7440(1993).
CC -1- FUNCTION: METHYLTRANSFERASE THAT CATALYZES THE METHYLATION OF
CC C-17 IN PRECORRIN-3B TO FORM PRECORRIN-4.
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PRECORRIN-3B = S-
CC ADENOSYL-L-HOMOCYSTEINE + PRECORRIN 4.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -1- SIMILARITY: TO S.TYPIHORIUM CBH; ALSO, LOW, TO OTHER
CC METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M59301; AAA25798.1; -.
CC PIR: E36145; E36145.
CC InterPro: IPR000878; Corrin_porph_methyltrnf.
CC Pfam: PF00590; TP_methylase; 1.
CC Cobalamin biosynthesis; Porphyrin biosynthesis; Transferrase;
CC Methyltransferase.
CC KW SEQUENCE 254 AA; 27105 MW; A30B027F4AFBE23F CRC64;
SO

```

```

alignment_scores:
  Quality: 49.50      Length: 34
  Ratio: 2.475       Gaps: 1
  Percent Similarity: 58.824   Percent Identity: 38.235

```

alignment\_block:

us-09-684-016-48411/rev x COBJ\_PSEDE ..

Align seg 1/1 to: COBJ\_PSEDE from: 1 to: 254

```

116 TGAATTAATTTGAAGATGATGACCTTACATGGAAGTACAGTCA 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
103 TrrysseValgluLeuValIleThrProglyValIleThrIleMetIleAla 119
66 GTGACGCCCTCGCATTT..GGCCATTAAGTCATCAGGGTTGTGAAGCAT 20
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 aValAlaAlaArgIleGluAlaProLeuGlyIleAspPheCysAlaIle 136
19 CT 18
||
136 er 136

```

seq\_name: SwissProt\_39:CAIF\_HUMAN

seq\_documentation\_block:

```

ID CAIF_HUMAN STANDARD; PRT; 1603 AA.
AC 007092;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)

```

```

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(XVI) CHAIN PRECURSOR.
GN COL16A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92335339; PubMed=1631157;
RA Pan T.C., Zhang R.Z., Mattei M.-G., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).
RN [2]
RP SEQUENCE OF 418-1603 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93203161; PubMed=1284248;
RA Yamaguchi N., Kimura S., McBride O.W., Hori H., Yamada Y.,
RA Kanamori T., Yamakoshi H., Nagai Y.;
RT "Molecular cloning and partial characterization of a novel collagen
RT chain, alpha 1(XVI), consisting of repetitive collagenous domains and
RL cysteine-containing non-collagenous segments.";
CC J. Biochem. 112:856-863(1992).
CC -1- FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE
CC THIS MOLECULE EITHER ELASTIC OR FLEXIBLE.
CC -1- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE
CC AMNION, A MEMBRANOUS TISSUE LINING THE AMNIOTIC CAVITY. WITHIN THE
CC AMNION, IT IS FOUND IN AN ACCELULAR, RELATIVELY DENSE LAYER OF A
CC COMPLEX NETWORK OF RETICULAR FIBERS. ALSO LOCATED TO A FIBROBLAST
CC LAYER BENEATH THIS DENSE LAYER. EXISTS IN TISSUES IN ASSOCIATION
CC WITH OTHER TYPES OF COLLAGEN.
CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING
CC GESTATION, AND DECREASE AT TERM.
CC -1- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE
CC TRIPLE-HELICAL DOMAINS (COL9 TO COL1) AND TEN NONTRIPLE-HELICAL
CC DOMAINS (NC10 TO NC1).
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF FIBRIL-ASSOCIATED COLLAGENS
CC WITH INTERRUPTED HELICES (FACIT).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M92642; AAA58427.1; -.
CC EMBL: S57132; AAB25797.1; -.
CC PIR: S23810; S23810.
CC MIM: 120326; -.
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR003129; TSPN.
CC Pfam: PF01391; Collagen; 16.
CC Pfam: PF02210; TSPN; 1.
CC SMART: SM00210; TSPN; 1.
CC KW Extracellular matrix; Connective tissue; Collagen; Hydroxylation;
CC Repeat; Signal.
CC FT SIGNAL 1 21
CC FT CHAIN 22 1603
CC FT DOMAIN 22 374
CC FT DOMAIN 375 505
CC FT DOMAIN 506 520
CC FT DOMAIN 521 554
CC FT DOMAIN 555 571
CC FT DOMAIN 572 630
CC FT DOMAIN 631 651
CC FT DOMAIN 652 722

```







CC -1- CATALYTIC ACTIVITY: EXOHYDROLYSIS OF THE NON-REDUCING TERMINAL  
 CC GLUCOSE RESIDUE IN THE MANNOSE-OLIGOSACCHARIDE  
 CC GIC(3)MAN(9)GICNAC(2).  
 CC -1- PATHWAY: FIRST ENZYME IN THE N-LINKED OLIGOSACCHARIDE PROCESSING  
 CC PATHWAY.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, ENDOPLASMIC  
 CC RETICULUM.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 63 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; X87237; CAA60683.1; .  
 CC DR MIM; 601336; .  
 CC KW Hydrolase; Glycosidase; Glycoprotein; Transmembrane; Signal-anchor;  
 CC Endoplasmic reticulum.  
 CC FT DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 39 59 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC FT DOMAIN 60 836 LUMENAL (POTENTIAL).  
 CC CAROHD 657 657 N-LINKED (GLCNAc...).  
 CC FT SEQUENCE 836 AA; 91840 MW; 57CE4A82126004AC CRC64;

alignment\_scores:  
 Quality: 49.00 Length: 42  
 Ratio: 2.130 Gaps: 1  
 Percent Similarity: 54.762 Percent Identity: 28.571

alignment\_block:  
 US-09-684-016-48411/rev x GCS1\_HUMAN ..

Align seg 1/1 to: GCS1\_HUMAN from: 1 to: 836

```

122 GTTGTGGAATTAATGAGATACACCTTACAT..... 84
      |||  |||  |||  |||  |||  |||  |||  |||
63 LeuAlaIrrPyrrAlaArgAlaValThrLeuHisSerAlaProPr 79
      |||  |||  |||  |||  |||  |||  |||  |||
83 .GTGAACCTTACCAAGTCAGCCCTGCATTGGCGCTTAAAGTATC 35
      |||  |||  |||  |||  |||  |||  |||  |||
79 oValLeuProAlaIspSerSerProAlaValAlaProAspLeuPheT 96

```

seq\_name: SwissProt\_39:TFEB\_MOUSE

seq\_documentation\_block:  
 ID TFEB\_MOUSE STANDARD: PRT; 475 AA.  
 AC O9R210:  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TRANSCRIPTION FACTOR EB.  
 GN TFEB OR TCPEB.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=BALB/C.  
 RX MEDLINE=99156662; PubMed=10036191.  
 RA Rehl M., Den Elzen N., Cassidy A.I., Ostrowski M.C., Hume D.A.;  
 RT "Cloning and characterization of the murine genes for bHLH-zip  
 RT transcription factors TFEB and TFEB reveal a common gene organization  
 RT for all M17 subfamily members.";

RL Genomics 56:111-120(1999).  
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR BINDS TO THE USP/MLTF SITE  
 CC AND PROBABLY RECOGNIZES E-BOX SEQUENCES IN THE HEAVY-CHAIN  
 CC IMMUNOGLOBULIN ENHANCER (BY SIMILARITY).  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; AF079095; AAD20979.1; ALT\_INIT.  
 CC DR MGD; MGI:103270; Tcfef.  
 CC DR InterPro; IPR001092; HLH\_dlm.  
 CC DR InterPro; IPR003015; HLH\_Myc.  
 CC DR Pfam; PF00010; HLH\_1.  
 CC DR SMART; SM00353; HLH; 1.  
 CC DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
 CC KW Transcription regulation; DNA-binding; Nuclear protein.  
 CC FT DOMAIN 10 43 GLN-RICH  
 CC FT DOMAIN 15 164 STRONG TRANSCRIPTION ACTIVATION DOMAIN  
 CC FT DNA\_BIND 232 247 BASIC DOMAIN.  
 CC FT DOMAIN 248 288 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 CC FT DOMAIN 297 318 LECICINE-ZIPPER (POTENTIAL).  
 CC FT DOMAIN 363 413 PRO-RICH.  
 CC FT DOMAIN 35 43 POLY-GLN.  
 CC FT SEQUENCE 475 AA; 52614 MW; 4249C6F501968326 CRC64;

alignment\_scores:  
 Quality: 48.50 Length: 28  
 Ratio: 2.553 Gaps: 1  
 Percent Similarity: 67.857 Percent Identity: 39.286

alignment\_block:  
 US-09-684-016-48411/rev x TFEB\_MOUSE ..

Align seg 1/1 to: TFEB\_MOUSE from: 1 to: 475

```

86 CATGTGAACCTTACCAAGTCAGCCCTGCATTGGCGCTTAAAGTCA 37
      |||  |||  |||  |||  |||  |||  |||  |||
106 HisValser...ProAlaGlnGlySerProAlaProAlaAlaIase 121
      |||  |||  |||  |||  |||  |||  |||  |||
36 TCAGGTTGTAGGCAATTCACATTTACAAATACC 3
      |||  |||  |||  |||  |||  |||  |||  |||
121 rProGlyValArgAlaGlyHisValSerThr 132

```

seq\_name: SwissProt\_39:ZP2\_MOUSE

seq\_documentation\_block:  
 ID ZP2\_MOUSE STANDARD: PRT; 713 AA.  
 AC P20239;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ZONA PELLUCIDA Sperm-BINDING PROTEIN 2 PRECURSOR (ZONA PELLUCIDA  
 DE GLYCOPROTEIN ZP2) (ZONA PELLUCIDA PROTEIN A).  
 GN ZP2 OR ZP-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A., AND SEQUENCE OF 35-50 AND 428-437.  
 RP TISSUE=Ovary;







DR SMART: SM00423; PST: 1.  
 DR SMART: SM00327; VWA: 1.  
 DR PROSITE: PS00243; INTEGRIN\_BETA: 3.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_2.  
 KW Integrin, Cell adhesion; Receptor; Transmembrane; Glycoprotein;  
 Repeat; Extracellular matrix; Cytoskeleton; Signal; Disease mutation.  
 FT SIGNAL 1 22  
 FT CHAIN 23 769  
 FT DOMAIN 23 700  
 FT TRANSMEM 701 723  
 FT DOMAIN 724 769  
 FT DOMAIN 124 363  
 FT DOMAIN 449 617  
 FT REPEAT 449 496  
 FT REPEAT 497 540  
 FT REPEAT 541 581  
 FT REPEAT 582 617  
 FT DISULFID 25 447  
 FT DISULFID 33 43  
 FT DISULFID 36 73  
 FT DISULFID 46 62  
 FT DISULFID 191 198  
 FT DISULFID 246 286  
 FT DISULFID 386 400  
 FT DISULFID 420 662  
 FT DISULFID 445 449  
 FT DISULFID 459 470  
 FT DISULFID 467 506  
 FT DISULFID 472 481  
 FT DISULFID 483 497  
 FT DISULFID 512 517  
 FT DISULFID 514 549  
 FT DISULFID 519 534  
 FT DISULFID 536 541  
 FT DISULFID 537 562  
 FT DISULFID 559 590  
 FT DISULFID 564 573  
 FT DISULFID 575 582  
 FT DISULFID 596 601  
 FT DISULFID 598 643  
 FT DISULFID 603 612  
 FT DISULFID 615 618  
 FT DISULFID 622 631  
 FT DISULFID 628 631  
 FT DISULFID 647 670  
 FT SITE 397 399  
 FT CARBOHYD 50 50  
 FT CARBOHYD 116 116  
 FT CARBOHYD 254 254  
 FT CARBOHYD 501 501  
 FT CARBOHYD 642 642  
 FT VARIANT 128 128  
 SQ SEQUENCE 769 AA: 84400 MW: 5903ADP4E8998CEA CRC64;

alignment\_scores:  
 Quality: 48.50 Length: 19  
 Ratio: 3.731 Gaps: 1  
 Percent Similarity: 68.421 Percent Identity: 42.105

alignment\_block:  
 US-09-684-016-48411 x ITB2\_BOVIN

Align seg 1/1 to: ITB2\_BOVIN from: 1 to: 769

8 TGTAAATGAGATGCTTACACCCGTGATGCGCAATGCGA 57  
 |||::|||::|||::|||::|||::|||::|||::|||::|||  
 601 CysGlnCysAspProGlyTyrGlnPro...ProLeuCysserGlyCysPpr 616  
 58 GGCGTCG 64  
 |||::|||::|||::|||::|||::|||::|||::|||  
 616 OGlyCys 618

seq\_name: SwissProt\_39:YCZ6\_YEAST  
 seq\_documentation\_block:  
 ID YCZ6\_YEAST STANDARD; PRT; 832 AA.  
 AC P25611;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PUTATIVE 95.7 KDa TRANSCRIPTIONAL REGULATORY PROTEIN IN PAU3-AAD3  
 DE INTERGENIC REGION.  
 GN YCR106W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_Taxid=4932;  
 RN [1]  
 RP SEQUENCE OF 1-517 FROM N.A.  
 RA van der Linden C.G., Maurer C.T.C., Planta R.J.,  
 RA van Vliet-Reedijk J.C.;  
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 130-832 FROM N.A.  
 RA Gresson M., Jauniaux J.-C., Urrestarazu L.A.;  
 RA Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP PRESENCE OF A 2N(2)-CYS(6) FUNGAL-TYPE BINUCLEAR CLUSTER.  
 RX MEDLINE=92350247; PubMed=1641000;  
 RA Bork P., Ouzounis C., Sander C., Scharf M., Schneider R.,  
 RA Sonnhammer E.;  
 RT "What's in a genome?";  
 RL Nature 358:287-287(1992).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- SIMILARITY: CONTAINS 1 2N(2)-CYS(6), FUNGAL-TYPE BINUCLEAR  
 CC CLUSTER DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X59720; CAA42238.1; -;  
 DR PIR: S19418; S19418.  
 DR HSSP: P12351; 1PYC.  
 DR SGD: S0000703; YCR106W.  
 DR InterPro: IPR001138; Zn2\_Cy6\_fungal.  
 DR Pfam: PF001172; Zn\_Clus: 1.  
 DR SMART: SM00066; GAL4; 1.  
 DR PROSITE: PS00463; ZN2\_Cy6\_FUNGAL\_1; 1.  
 DR PROSITE: PS00048; ZN2\_Cy6\_FUNGAL\_2; 1.  
 KW Hypothetical protein; Transcription regulation; DNA-binding;  
 KW Nuclear protein; Zinc; Metal-binding.  
 FT DNA\_BIND 15 42  
 FT SEQUENCE 832 AA: 95689 MW: 7DF91D86AF08E4B3 CRC64;

alignment\_scores:  
 Quality: 48.50 Length: 21  
 Ratio: 3.233 Gaps: 1  
 Percent Similarity: 71.429 Percent Identity: 47.619

alignment\_block:  
 US-09-684-016-48411/rev x YCZ6\_YEAST

Align seg 1/1 to: YCZ6\_YEAST from: 1 to: 832

122 GTTGTTGGAATTAATGAGATAGTACCTTACATGGAACCTTAC 73  
 ::|||::|||::|||::|||::|||::|||::|||::|||  
 457 ILeValTrpaenLysTyrLysPheHisValIleLeuSer...LysLeuMe 472







Quality: 48.00 Length: 27  
Ratio: 3.429 Gaps: 1  
Percent Similarity: 51.852 Percent Identity: 33.333

alignment\_block:

US-09-684-016-48411 x DMSB\_HAEIN ..

Align seg 1/1 to: DMSB\_HAEIN from: 1 to: 205

2 AGCATGTGTAATGTGAGATGCCCTTACACCCCTATGACCTA..... 43

103 ArgTyrCysHisMetAlaCysProTyrAspAlaProGlnTyrAspAlaG1 119

44 .....ATGGCCATCGAGGATCG 64

119 nlysglyHisMetThrTyrCysAspGlyCys 129

seq\_name: SwissProt\_39:VIPR\_MELGA

seq\_documentation\_block:

ID VIPR\_MELGA STANDARD; PRT; 260 AA.

AC 091085; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE VASOACTIVE INTESTINAL POLYPEPTIDE RECEPTOR (VIP-R) (VIP RECEPTOR)

DE (FRAGMENT)

OS Meleagris gallopavo (Common turkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.

NCBI\_TaxID=9103;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Small intestine;

RX MEDLINE=96206340; PubMed=8618952;

RA Xu M., Proudman J.A., Pits G.R., Wong E.A., Foster D.N.,

el Halawani M.E.; "Vasoreactive intestinal peptide stimulates prolactin mRNA expression

in turkey pituitary cells: effects of dopaminergic drugs.";

Proc. Soc. Exp. Biol. Med. 212:52-62(1996).

-1- FUNCTION: THIS IS A RECEPTOR FOR VIP. THE ACTIVITY OF THIS

RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL

CYCLASE.

-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: U31991; AAA99740.1; -

DR GCRB; GCR\_1749; -

DR InterPro: IPR000832; GPCR\_secretin.

DR Pfam: PF00002; 7tm\_2; 1

DR PROSITE: PS00649; G\_PROTEIN\_RECEP\_F2\_1; PARTIAL.

DR PROSITE: PS00650; G\_PROTEIN\_RECEP\_F2\_2; 1

KW G-protein coupled receptor; Transmembrane; Glycoprotein.

FT NON\_TER

FT DOMAIN

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

FT TRANSMEM

SEQUENCE 260 AA; 30358 MW; 15761E6AB5B23D5A CRC64;

alignment\_scores:

Quality: 48.00 Length: 23

Ratio: 3.000 Gaps: 1

Percent Similarity: 69.565 Percent Identity: 39.130

alignment\_block:

US-09-684-016-48411/rev x VIPR\_MELGA ..

Align seg 1/1 to: VIPR\_MELGA from: 1 to: 260

75 TACAGTCAGTGCAGACCCCTGCATTCGCGCATTTAGTCATCAGGTTGNA 26

2 PheGlnSerGlyGlnProGlnHisCysPheValSerSer.ValGlyCysL 18

25 AGGCATCTCACATTTC 9

18 ysAlaMetMetValPhe 23

seq\_name: SwissProt\_39:TFZ\_HUMAN

seq\_documentation\_block:

ID TFZ\_HUMAN STANDARD; PRT; 292 AA.

AC 016635; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE TAFAZZIN.

GN TAZ OR EFE2 OR G4.5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Skeletal muscle, and Heart;

RX MEDLINE=96224398; PubMed=8630491;

RA Blone S., D'Adamo P., Maestrini E., Gedeon A.K., Bolhuis P.A.,

Tonolo D.;

"A novel X-linked gene, G4.5, is responsible for Barth syndrome.";

Nat. Genet. 12:385-389(1996).

-1- SUBCELLULAR LOCATION: ISOFORMS WITH HYDROPHOBIC N-TERMINUS ARE

THOUGHT TO BE MEMBRANE ANCHORED. SHORTEST FORMS, LACKING THE

HYDROPHOBIC STRETCH, MAY BE SOLUBLE CYTOPLASMIC PROTEINS.

-1- ALTERNATIVE PRODUCTS: UP TO 10 ISOFORMS. SHORTER TAFAZZINS START

WITH THE SECOND METHIONINE AT POS. 25. ALTERNATIVE SPLICING OF THE

HYDROPHILIC CENTRAL REGION (EXONS 5-7) MAY PRODUCE 5 VARIANTS. TWO

ISOFORMS, CONTAINING ALL THREE EXONS OR LACKING EXON 5 ONLY ARE

CONSTANTLY ABUNDANT. THE SEQUENCE SHOWN HERE IS THAT OF THE

LONGEST ISOFORM.

-1- TISSUE SPECIFICITY: HIGH LEVELS IN CARDIAC AND SKELETAL MUSCLE. UP

TO 10 ISOFORMS CAN BE PRESENT IN DIFFERENT AMOUNTS IN DIFFERENT

TISSUES. MOST ISOFORMS ARE UBIQUITOUS. ISOFORMS THAT LACK THE N-

TERMINUS ARE FOUND IN LEUKOCYTES AND FIBROBLASTS, BUT NOT IN HEART

AND SKELETAL MUSCLE. SOME FORMS APPEAR RESTRICTED TO CARDIAC AND

SKELETAL MUSCLE OR TO LEUKOCYTES.

-1- DOMAIN: THE HYDROPHILIC DOMAIN MAY SERVE AS AN EXPOSED LOOP

INTERACTING WITH OTHER PROTEINS.

-1- DISEASE: DEFECTS IN TAZ ARE THE CAUSE OF BARTH SYNDROME (BTHS), A

SEVERE INHERITED DISORDER, OFTEN FATAL IN CHILDHOOD, CHARACTERIZED

BY CARDIAC AND SKELETAL MYOPATHY, SHORT STATURE AND NEUTROPENIA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: X92763; CAA63419.1; -











[illegible]

RT "Regulatory elements and transcriptional regulation by testosterone  
RA Metis M., Tilmusk T., Allikmeets R., Saarma M., Pärsson H.;  
RX MEDLINE=93077038; PubMed=1446821;  
RP TISSUE=Liver;  
RP SEQUENCE OF 1-22 FROM N.A.  
RP 1421



RT and retinoic acid of the rat nerve growth factor receptor promoter.";  
 RL Gene 121:247-254(1992).  
 [3]  
 RP STRUCTURE BY NMR OF 334-418.  
 RX MEDLINE-97449145; PubMed-9305641;  
 RA Liepinsh E., Ilag L., Otting G., Ibanez C.F.;  
 RT "NMR structure of the death domain of the p75 neurotrophin receptor.";  
 RL EMBL J. 16:4999-5005(1997).  
 CC -1- FUNCTION: LOW AFFINITY RECEPTOR WHICH CAN BIND TO NGF, BDNF,  
 CC NT-3, AND NT-4.  
 CC -1- SUBUNIT: NGF RECEPTOR CAN FORM A HOMODIMER THROUGH DISULFIDE  
 CC BOND FORMATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- PTM: N- AND O-GLYCOSYLATED AND IS PHOSPHORYLATED ON SERINE.  
 CC -1- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.  
 CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X05137; CAA28783.1; -  
 CC EMBL: X61269; -; NOT\_ANNOTATED\_CDS.  
 CC PIR: A26431; A26431.  
 CC PDB: INGR; 29-70U-97.  
 CC InterPro: IPR000488; Death.  
 CC InterPro: IPR001368; TNFR\_c6.  
 CC Pfam: PF00531; death; 1.  
 CC Pfam: PF00020; TNFR\_c6; 4.  
 CC SMART: SM00005; DEATH; 1.  
 CC SMART: SM00208; TNFR; 3.  
 CC PROSITE: PS00652; TNFR\_NGFR\_1; 3.  
 CC PROSITE: PS50050; TNFR\_NGFR\_2; 4.  
 CC PROSITE: PS50017; DEATH\_DOMAIN; 1.  
 CC Receptor; Neurogenesis; Transmembrane; Glycoprotein; Repeat;  
 CC Phosphorylation; Signal; 3D-structure.  
 CC SIGNAL 1 29  
 CC CHAIN 30 425 LOW-AFFINITY NERVE GROWTH FACTOR  
 CC RECEPTOR.  
 CC EXTRACELLULAR (POTENTIAL).  
 CC POTENTIAL.  
 CC CYTOPLASMIC (POTENTIAL).  
 CC 4 X TNFR-CYS.  
 CC TNFR-CYS 1.  
 CC REPEAT 67 108 TNFR-CYS 2.  
 CC REPEAT 109 148 TNFR-CYS 3.  
 CC REPEAT 149 190 TNFR-CYS 4.  
 CC DOMAIN 198 249 SER/THR-RICH.  
 CC 419 DEATH.  
 CC DISULFID 334 419 BY SIMILARITY.  
 CC DISULFID 33 44 BY SIMILARITY.  
 CC DISULFID 45 58 BY SIMILARITY.  
 CC DISULFID 48 65 BY SIMILARITY.  
 CC DISULFID 68 84 BY SIMILARITY.  
 CC DISULFID 87 100 BY SIMILARITY.  
 CC DISULFID 90 108 BY SIMILARITY.  
 CC DISULFID 110 123 BY SIMILARITY.  
 CC DISULFID 126 139 BY SIMILARITY.  
 CC DISULFID 129 147 BY SIMILARITY.  
 CC DISULFID 150 165 BY SIMILARITY.  
 CC DISULFID 168 181 BY SIMILARITY.  
 CC DISULFID 171 189 BY SIMILARITY.  
 CC CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC SEQUENCE 425 AA; 45432 MW; B2E15D94D3827F8 CRC64;

alignment\_scores:  
 Quality: 48.00  
 Ratio: 3.692

Length: 20  
 Gaps: 0

Percent Similarity: 65.000 Percent Identity: 35.000  
 alignment\_block:  
 US-09-684-016-48411 x NGFR\_RAT ..  
 Align seq 1/1 to: NGFR\_RAT from: 1 to: 425  
 8 TGTAAATGTGAGATGCGCTTACACCGTCGATGACCTAATGGCGCAATGCGA 57  
 |||:||||| |||:|||||  
 108 GysarGcysAlaTyrGlyTyrTyrGlnAspGluGlnThrGlyHisCysGsl 124  
 58 GGCGTGCACCT 67  
 |||:|||||  
 124 uAlacysSer 127  
 seq\_name: SwissProt\_39:GAC3\_HUMAN  
 seq\_documentation\_block:  
 ID GAC3\_HUMAN STANDARD; PRT; 467 AA.  
 AC 099928; Q9HD46; Q9NYT2;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR GAMMA-3 SUBUNIT PRECURSOR (GABA(A)  
 DE RECEPTOR).  
 GN GABRG3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Fetal brain;  
 RX MEDLINE-96360042; PubMed-8719414;  
 RA Hadingham K.L., Walford K.A., Thompson S.A., Palmer K.J.,  
 RA Whiting P.J.;  
 RT "Expression and pharmacology of human GABA receptors containing  
 RT gamma 3 subunits".  
 RL Eur. J. Pharmacol. 291:301-309(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Han M.K., Chen Y.-H., Nurni E.L., Sutcliffe J.S.;  
 RT "Gene structure and organization of the GABA receptor gamma 3 subunit  
 RT (GABRG3) gene".  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 19-467 FROM N.A.  
 RA Kim S.-Y., Gonen D., Yang Z.-Y., Reliford A., Dy A., Leventhal B.L.,  
 RA Cook E.H., Jr.;  
 RT "Genomic organization of GABA receptor gamma 3 subunit gene  
 RT (GABRG3)".  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE  
 CC VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE  
 CC GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE  
 CC CHANNEL.  
 CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)  
 CC RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- MISCELLANEOUS: THIS SUBUNIT CARRIES THE BENZODIAZEPINE BINDING  
 CC SITE.  
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: S82769; AAB39369.1; -  
 CC EMBL: AF269144; AAF99696.1; -



GN MML / UK M14MMH  
OS Homo sapiens (H)

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Carnivora; Homidae; Homo.  
CC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RC TISSUE=Monocytic Leukemia;  
RX MEDLINE=99402951; PubMed=10471807;  
RA Kajita M., Kinoh H., Ito N., Takamura A., Itoh Y., Okada A., Sato H.,  
RT Seiki M.;  
RT "Human membrane type-4 matrix metalloproteinase (MT4-MMP) is encoded  
RT by a novel major transcript: isolation of complementary DNA clones  
RT for human and mouse mt4-mmp transcripts.";  
RL FEBS Lett. 457:353-356(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=96234364; PubMed=8640782;  
RA Puente X.S., Pendas A.M., Llano E., Velasco G., Lopez-Otin C.;  
RT "Molecular cloning of a novel membrane-type matrix metalloproteinase  
RT from a human breast carcinoma.";  
RL Cancer Res. 56:944-949(1996).  
RN [3]  
RP SEQUENCE OF 129-302 FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=20020281; PubMed=10551873;  
RA Wang Y., Johnson A.R., Ye Q.-Z., Dyer R.D.;  
RT "Catalytic activities and substrate specificity of the human membrane  
RT type 4 matrix metalloproteinase catalytic domain.";  
RL J. Biol. Chem. 274:33043-33049(1999).  
RN [4]  
RP GPI-ANCHOR.  
RX MEDLINE=20036570; PubMed=10567400;  
RA Itoh Y., Kajita M., Kinoh H., Mori H., Okada A., Seiki M.;  
RT "Membrane type 4 matrix metalloproteinase (MT4-MMP, MMP-17) is a  
RT glycosylphosphatidylinositol-anchored proteinase.";  
RL J. Biol. Chem. 274:34260-34266(1999).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=20008793; PubMed=10543448;  
RA Kolkebrock H., Essers L., Ubrich N., Will H.;  
RT "Biochemical characterization of the catalytic domain of membrane-type  
RT 4 matrix metalloproteinase.";  
RL J. Biol. Chem. 274:1103-1108(1999).  
CC -1- FUNCTION: ENDOPEPTIDASE THAT DEGRADES VARIOUS COMPONENTS OF THE  
CC EXTRACELLULAR MATRIX, SUCH AS FIBRIN. MAY BE INVOLVED IN THE  
CC ACTIVATION OF MEMBRANE-BOUND PRECURSORS OF GROWTH FACTORS OR  
CC INFLAMMATORY MEDIATORS, SUCH AS TUMOR NECROSIS FACTOR-ALPHA. MAY  
CC ALSO BE INVOLVED IN TUMORAL PROCESS. NOT OBVIOUS IF ABLE TO  
CC PROTEOLYTICALLY ACTIVATES PROGELATINASE A. DOES NOT HYDROLYSE  
CC COLLAGEN TYPES I, II, III, IV AND V, GELATIN, FIBRONECTIN,  
CC LAMININ, DECORIN NOR ALPHA1-ANTITRYPSIN.  
CC -1- CATALYTIC ACTIVITY: CLEAVES PRO-TNF-ALPHA AT THE 74-ALA-|-GLN-75  
CC SITE.  
CC -1- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC SHORT/PUNTE-TYPE FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LEUKOCYTES, COLON, OVARY  
CC TESTIS AND BREAST CANCER. EXPRESSED ALSO IN MANY TRANSFORMED AND  
CC NON-TRANSFORMED CELL TYPES.  
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUZIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC  
CC METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----



RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=95220383; PubMed=7705369;  
R1 Yamasu K., Watanabe H., Koyuchi C., Soma G.-I., Mizuno D.-I.,  
RA Aasaka K., Shimada H., Suyehitsu T., Ishihara K.;  
RT Molecular cloning of a cDNA that encodes the precursor to several  
RT exoglycosidase-inducing peptides, epidermal growth factor-related



RT polypeptides of the sea urchin Anthocidaris crassispina.";  
 RL Eur. J. Biochem. 228:515-523(1995).  
 RN [2]  
 RP SEQUENCE OF 106-158 AND 179-230.  
 RC TISSUE=Embryo;  
 RX MEDLINE=89230051; PubMed=2713739;  
 RA Suyemitsu T., Asami-Yoshizumi T., Noguchi S., Tonegawa Y.,  
 RA Ishihara K.;  
 RT "The exogastrula-inducing peptides in embryos of the sea urchin,  
 RT Anthocidaris crassispina -- isolation and determination of the primary  
 RT structure.";  
 RL Cell Differ. Dev. 26:53-66(1989).  
 RN [3]  
 RP SEQUENCE OF 47-104.  
 RC TISSUE=Embryo;  
 RX MEDLINE=90028385; PubMed=2804137;  
 RA Suyemitsu T., Tonegawa Y., Ishihara K.;  
 RT "Amino acid sequence of exogastrula-inducing peptide C from the sea  
 RT urchin, Anthocidaris crassispina.";  
 RL Biochim. Biophys. Acta 999:24-28(1989).  
 CC -1- FUNCTION: THE EGIP PEPTIDES ARE FACTORS EFFECTIVE TO EXTRUDE THE  
 CC ARCHENTERON TOWARD OUTSIDE OF EMBRYOS. MAY HAVE A ROLE IN THE  
 CC INDUCTION OF GASTRULATION.  
 CC -1- SIMILARITY: EACH OF THE 4 PEPTIDES CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z48184; CA88234.1; -;  
 DR PIR; A48167; A48167.  
 DR PIR; A48167; A48167.  
 DR PIR; S06637; S06637.  
 DR HSSP; P00592; JP2P.  
 DR InterPro; IPR000561; EGF-like.  
 DR Pfam; PF00008; EGF; 2.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00001; EGF-like; 2.  
 DR PROSITE; PS00186; EGF\_1; 4.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 KW Signal; Developmental protein; EGF-like domain; Extracellular matrix;  
 KW Repeat; Gastrulation.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT PEPTIDE 47 104 EXOGASTRULA-INDUCING PEPTIDE C (EGIP-C).  
 FT PEPTIDE 106 158 EXOGASTRULA-INDUCING PEPTIDE D (EGIP-D).  
 FT PEPTIDE 179 230 EXOGASTRULA-INDUCING PEPTIDE A (EGIP-A).  
 FT PEPTIDE 251 310 EGIP-X (PROBABLE).  
 FT DOMAIN 48 91 EGF-LIKE 1.  
 FT DOMAIN 107 154 EGF-LIKE 2.  
 FT DOMAIN 180 226 EGF-LIKE 3.  
 FT DOMAIN 252 298 EGF-LIKE 4.  
 FT DISULFID 52 65 BY SIMILARITY.  
 FT DISULFID 59 75 BY SIMILARITY.  
 FT DISULFID 77 90 BY SIMILARITY.  
 FT DISULFID 111 124 BY SIMILARITY.  
 FT DISULFID 118 138 BY SIMILARITY.  
 FT DISULFID 140 153 BY SIMILARITY.  
 FT DISULFID 184 197 BY SIMILARITY.  
 FT DISULFID 191 211 BY SIMILARITY.  
 FT DISULFID 213 225 BY SIMILARITY.  
 FT DISULFID 256 269 BY SIMILARITY.  
 FT DISULFID 263 283 BY SIMILARITY.  
 FT DISULFID 285 297 BY SIMILARITY.  
 SQ SEQUENCE 325 AA; 36462 MW; 1FDD2577B3617306A CRC64;

Percent Similarity: 35.897 Percent Identity: 25.641

alignment\_block:

US-09-684-016-48411 x EGIP\_ANTCR ..

Align seg 1/1 to: EGIP\_ANTCR from: 1 to: 325

5 TATGTAAATGTGAGATGCTTAC..... 28  
 |||||.....  
 74 TyrcyslscysThrleuProTyArgvalGlySerclusercercy 90  
 29 .....AACCCGATGACC 41  
 90 sTyrcMetProLysaspLysgluLaspvalGluIleGluThrLysaspr 107  
 42 TAAAGCGCAATCGAG 58  
 :::::.....  
 107 hrValAlaArgCysGlu 112

alignment\_scores:

Quality: 47.50

Length: 39

Ratio: 3.393

Gaps: 1







BE BLANK (USPTO)

THIS PAGE BLANK (USPTO)



OM of: US-09-684-016-48411 to: SPTRMBL\_17:\* out\_format : pfs  
Date: Mar 19, 2002 9:34 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

```
-MODE=frame:r2p.model -DEV=x1h  
-Q=/cgcn2.1/USPTO_sptool/US09684016/runatc.19032002.084440.20931/app-query.fasta.1.182  
-DB=SPTRMBL_17 -QFMT=fastan -SUFFIX=std.rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FCAPOP=6.000 -FCAPEXT=7.000 -YCAPOP=10.000 -YCAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human:0.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALLEN=45 -MODE=LOCAL -OUTFMT=pfs  
-NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09684016.cgcn1.1.96 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPY -WAIT -THREADS=1
```

## Search information block:

Query: US-09-684-016-48411  
Query length: 123  
Database: SPTRMBL\_17:\*  
Database sequences: 473505  
Database length: 146272329  
Search time (sec): 65.740000

## score\_list:

Sequence	Strd	Orig	ZScore	Escore	len	Documentation
SP.plant:065462	+	138.00	374.72	6.7e-13	201	065462 arabidopsis thaliana (mc
SP.plant:043479	+	127.00	342.84	3.5e-11	227	043479 hordeum vulgare (barley)
SP.plant:09F9E9	+	124.00	333.34	1.0e-10	228	09F9E9 arabidopsis thaliana (mc
SP.plant:09SV14	+	124.00	334.51	1.0e-10	258	09SV14 arabidopsis thaliana (mc
SP.plant:09XJ51	+	98.00	268.78	1.5e-06	71	09XJ51 triticum aestivum (wheat)
SP.rodent:09WTU0	+	77.00	191.06	0.0021	1096	09WTU0 mus musculus (mouse)
SP.human:09Y6M4	+	77.00	191.04	0.0021	1099	09Y6M4 homo sapiens (human)
SP.invertebrate:09VUE3	+	76.50	196.37	0.0029	404	09VUE3 drosophila melanogaster
SP.human:09H4H2	+	74.00	184.97	0.0065	790	09H4H2 homo sapiens (human)
SP.human:09U0P1	+	74.00	182.66	0.0062	1084	09U0P1 homo sapiens (human)
SP.fungi:003012	+	66.50	169.04	0.1089	353	003012 saccharomyces cerevisiae
SP.fungi:074508	+	62.50	156.51	0.4521	424	074508 schizosaccharomyces pombe
SP.invertebrate:045410	+	62.00	155.41	0.5453	405	045410 caenorhabditis elegans
SP.invertebrate:09N1C1	+	62.00	153.86	0.5284	451	09N1C1 caenorhabditis elegans
SP.invertebrate:045407	+	60.00	136.79	0.8652	2781	045407 caenorhabditis elegans
SP.human:09U1G2	+	60.00	132.90	0.7996	4957	09U1G2 homo sapiens (human)
SP.human:014687	+	60.00	132.49	0.7931	5262	014687 homo sapiens (human)
SP.invertebrate:022662	+	59.50	148.17	1.134	416	022662 caenorhabditis elegans
SP.plant:09T0A1	+	59.50	147.75	1.133	443	09T0A1 arabidopsis thaliana (mc
SP.invertebrate:09VRV6	+	57.50	139.39	2.260	662	09VRV6 drosophila melanogaster
SP.fungi:09U779	+	57.00	132.09	2.77	1588	09U779 schizosaccharomyces pom
SP.human:09BQW0	+	56.50	140.33	4.03	379	09BQW0 homo sapiens (human)
SP.human:09BM03	+	56.50	137.89	3.84	544	09BM03 homo sapiens (human)
SP.human:09N1M8	+	56.50	137.67	3.82	562	09N1M8 homo sapiens (human)
SP.human:09B7C0	+	56.50	137.67	3.82	562	09B7C0 homo sapiens (human)
SP.rodent:09WVU0	+	56.50	137.08	3.78	614	09WVU0 mus musculus (mouse)
SP.invertebrate:09W664	+	56.50	136.79	3.75	641	09W664 barbus barbus (barbel)
SP.human:015043	+	56.50	133.85	3.54	991	015043 homo sapiens (human)
SP.human:09H4G6	+	56.50	133.52	3.51	1042	09H4G6 homo sapiens (human)
SP.archaea:09Y8L2	+	56.00	136.48	4.60	544	09Y8L2 aeropyrum pernix (hyoch
SP.bacteria:09R2L4	+	55.50	137.79	5.83	363	09R2L4 legionella pneumophila
SP.human:09NR48	+	55.50	123.64	4.37	2969	09NR48 homo sapiens (human)
SP.human:09H0S7	+	55.00	141.11	7.68	180	09H0S7 homo sapiens (human)
SP.invertebrate:09XMD7	+	55.00	136.29	6.97	368	09XMD7 caenorhabditis elegans
SP.plant:09FN15	+	55.00	126.76	5.75	1516	09FN15 arabidopsis thaliana (m
SP.invertebrate:09U0P3	+	54.50	124.73	6.81	1645	09U0P3 caenorhabditis elegans
SP.invertebrate:044498	+	54.50	124.75	6.80	1655	044498 caenorhabditis elegans
SP.invertebrate:09VR08	+	54.50	119.37	6.10	3680	09VR08 drosophila melanogaster
SP.invertebrate:09V7S1	+	54.00	129.90	9.31	625	09V7S1 drosophila melanogaster

SP.invertebrate:024459 + 54.00 127.77 8.92 857 024459 drosophila melanog  
SP.invertebrate:09V8C2 + 54.00 126.46 8.69 1042 09V8C2 drosophila melano  
SP.plant:09F9E5 + 53.50 129.11 11.30 570 09F9E5 arabidopsis thaliana  
SP.invertebrate:09V678 + 53.50 120.60 9.52 2016 09V678 drosophila melano  
SP.virus:098VX5 + 53.00 134.82 15.65 198 098VX5 human immunodeficien

seq\_name: sp\_plant:065462

seq\_documentation\_block:  
ID 065462 PRELIMINARY; PRT: 201 AA.  
AC 065462;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE RECEPTOR LIKE PROTEIN (FRAGMENT).  
GN FIN20.240 OR ATG622140.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Wedler H., Wambutt R., Bancroft I., Mewes H.W., Mayer K.,  
RA Schueller C.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-0 FROM N.A.  
RA Robben J., Gymnompres B., Volckaert G., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL022140; CAA18117.1; -  
DR EMBL; AL161556; CAB7169.1; -  
DR Mendel; 29749; Arabid; 2000; 29749.  
DR InterPro; IPR001025; BAH.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001965; PHD.  
DR Pfam; PF01426; BAH; 1.  
DR Pfam; PF00628; PHD; 1.  
DR SMART; SM00439; BAH; 1.  
DR SMART; SM00249; PHD; 1.  
DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.  
SO SEQUENCE 201 AA; 22579 MW; 4C48904BB9E61498 CMC64;

## alignment\_scores:

Quality: 138.00 Length: 23  
Ratio: 6.273 Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 91.304

## alignment\_block:

US-09-684-016-48411 x 065462 ..  
Align seg 1/1 to: 065462 from: 1 to: 201

5 TATTGTAATGTCAGATGCTTACCAACCTGATGACCTAATGCGCATG 54  
|||||  
115 TTYCTSLYSCYSGLMEPTOTYRASNPROASPLeuMetValGlnCys 131  
55 CGAGGCGTGCAGCTGACTG 73  
|||||  
131 sGluLysCysAspTyr 137  
seq\_name: sp\_plant:043479



```

seq_documentation_block:
ID      Q43479      PRELIMINARY;      PRT;      227 AA.
AC      Q43479;
DT      01-NOV-1996 (TREMblrel. 01, Created)
DT      01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE      (DMS576).
GN      ES43.
OS      Hordeum vulgare (Barley).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC      Triticaceae; Hordeum.
OX      NCBI_TaxID=4513;
RN
  [1]
  SEQUENCE FROM N.A.
RC      STRAIN=CV, DBS576; TISSUE=LEAF;
RA      Speulman E., Salamini F.;
RL      Plant Sci. 106:91-98(1995).
DR      EMBL; X77575; CA54682.1;
DR      Mendei.; 12800; Horvu; 2000; 12800.
DR      InterPro; IPR000561; EGF-like.
DR      InterPro; IPR001025; BAH.
DR      InterPro; IPR001965; PHD.
DR      Pfam; PF00628; PHD; 1.
DR      Pfam; PF01426; BAH; 1.
DR      PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR      SMART; SM00439; BAH; 1.
DR      SMART; SM00249; PHD; 1.
SQ      SEQUENCE 227 AA; 25741 MW; ECLAB6DB0A472E24 CRC64;

alignment_scores:
Quality: 127.00      Length: 23
Ratio: 6.048      Gaps: 0
Percent Similarity: 91.304      Percent Identity: 82.609

alignment_block:
US-09-684-016-48411 x Q43479 ..
Align seg 1/1 to: Q43479 from: 1 to: 227

5 TATTGTAATGTGAGATGCCTTACACCGCTGATGACCTAATGGCGAATG 54
:::|||||
147 PhcysLysCysGluMetProTyrAsnProAspAspleuMetValGlnC 163
||| |||
55 CGAGGCGTGCACCTGACTGG 73
||| |||
163 sgluGluCysSerAspTrrp 169

seq_name: sp_plant:Q9FEN9

seq_documentation_block:
ID      Q9FEN9      PRELIMINARY;      PRT;      228 AA.
AC      Q9FEN9;
DT      01-MAR-2001 (TREMblrel. 16, Created)
DT      01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE      PUTATIVE PHD FINGER TRANSCRIPTION FACTOR.
GN      SHL1.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_TaxID=3702;
RN
  [1]
  SEQUENCE FROM N.A.
RA      Oh M., Torisky R.S., Braam J., Altmann T., Clouse S.D.;
RT      "PHD Finger Dependent Binding of SHL1 to a Specific Promoter Region of
  Arabidopsis TCH4 in vitro.";
RL      Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AF277453; AAG21353.1;
DR      InterPro; IPR001025; BAH; 1.
DR      InterPro; IPR00164; Histone_H3.

```

```

DR      InterPro; IPR001965; PHD.
DR      Pfam; PF01426; BAH; 1.
DR      Pfam; PF00628; PHD; 1.
DR      SMART; SM00439; BAH; 1.
DR      SMART; SM00249; PHD; 1.
DR      PROSITE; PS00322; HISTONE_H3_1; UNKNOWN_1.
SQ      SEQUENCE 228 AA; 26122 MW; AB00BC81E5B8390A CRC64;

alignment_scores:
Quality: 124.00      Length: 23
Ratio: 5.636      Gaps: 0
Percent Similarity: 95.652      Percent Identity: 78.261

alignment_block:
US-09-684-016-48411 x Q9FEN9 ..
Align seg 1/1 to: Q9FEN9 from: 1 to: 228

5 TATTGTAATGTGAGATGCCTTACACCGCTGATGACCTAATGGCGAATG 54
:::|||||
141 PhcysLysCysGluMetProTyrAsnProAspAspleuMetValGlnC 157
||| |||
55 CGAGGCGTGCACCTGACTGG 73
||| |||
157 sgluGluCysSerAspTrrp 163

seq_name: sp_plant:Q9SVI4

seq_documentation_block:
ID      Q9SVI4      PRELIMINARY;      PRT;      258 AA.
AC      Q9SVI4;
DT      01-MAY-2000 (TREMblrel. 13, Created)
DT      01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE      ES43 LIKE PROTEIN (FRAGMENT).
GN      F19H22.200 OR AT4G639100.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_TaxID=3702;
RN
  [1]
  SEQUENCE FROM N.A.
RA      Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA      Mayer K.F.X., Scheller C.;
RL      Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN
  [2]
  SEQUENCE FROM N.A.
RA      EU Arabidopsis sequencing project;
RL      Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN
  [3]
  SEQUENCE FROM N.A.
RA      Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL      Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN
  [4]
  SEQUENCE FROM N.A.
RA      Zimmermann W., Grueneisen A., Wambutt R., Kalicki J., Wohlmann P.,
RA      Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL      Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN
  [5]
  SEQUENCE FROM N.A.
RA      EU Arabidopsis sequencing project;
RL      Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AL035679; CAB38830.1;
DR      EMBL; AL161594; CAB80573.1;
DR      InterPro; IPR001025; BAH.
DR      InterPro; IPR000561; EGF-like.
DR      InterPro; IPR000164; Histone_H3.
DR      InterPro; IPR001965; PHD.
DR      Pfam; PF01426; BAH; 1.
DR      Pfam; PF00628; PHD; 1.
DR      SMART; SM00439; BAH; 1.

```



DR SMART: SM00249; PHD; 1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE: PS00322; HISTONE\_H3\_1; UNKNOWN\_1.  
 FT NON\_TER 258  
 SO SEQUENCE 258 AA; 29675 MW; 17A2P292B1C1E77 CRC64;

alignment\_scores:  
 Quality: 124.00 Length: 23  
 Ratio: 5.636 Gaps: 0  
 Percent Similarity: 95.652 Percent Identity: 78.261

alignment\_block:  
 US-09-684-016-48411 x Q9SVI4 ..

Align seg 1/1 to: Q9SVI4 from: 1 to: 258

5 TATTGTAATGTGAGATGCTTACACCGTATGACCTAATGGCGCAATG 54  
 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 141 PhecysLysCysGluMetProTyrAsnProAspAspLeuMetValGlnCys 157  
 55 CGAGGGCTGCACCTGACTGG 73  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 157 sgluglncysLysSerGluTrp 163

seq\_name: sp\_plant:Q9XJ51

seq\_documentation\_block:  
 ID Q9XJ51 PRELIMINARY; PRT; 71 AA.

AC Q9XJ51;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ZINC-FINGER MOTIF (FRAGMENT).  
 GN WSR4.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CHINESE SPRING; TISSUE=ROOT;  
 RA Nemoto Y., Kawakami N., Sasakuma T.;  
 RT "Isolation of novel early salt-responding genes from wheat (Triticum  
 RT aestivum L.) by differential display."  
 RL Theor. Appl. Genet. 98:673-678(1999).  
 DR EMBL; AB011445; BAA82157.1; -.  
 DR InterPro: IPR001965; PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 KW ZINC.  
 FT NON\_TER 1  
 SO SEQUENCE 71 AA; 8056 MW; 9E06685BC184E423 CRC64;

alignment\_scores:  
 Quality: 98.00 Length: 17  
 Ratio: 6.125 Gaps: 0  
 Percent Similarity: 94.118 Percent Identity: 88.235

alignment\_block:  
 US-09-684-016-48411 x Q9XJ51 ..

Align seg 1/1 to: Q9XJ51 from: 1 to: 71

23 CCTTACACCTGATGACCTAATGCGCAATGCGAGGCTGCACCTGACTG 72  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1 ProTyrAsnProAspAspLeuMetValGlnCysglncysLysAspTr 17  
 73 G 73  
 1  
 17 p 17

seq\_name: sp\_rodent:Q9WTU0

seq\_documentation\_block:  
 ID Q9WTU0 PRELIMINARY; PRT; 1096 AA.

AC Q9WTU0;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE PHD-FINGER PROTEIN.  
 GN PHF2 OR GRC5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99160472; PubMed=10051327;  
 RA Hasenpusch-Theil K., Chadwick B.P., Theil T., Heath S.K.,  
 RA Wilkison D.G., Frischauf A.M.;  
 RT "PHF2, a novel PHD finger gene located on human chromosome 9q22."  
 RL Mamm. Genome 10:294-298(1999).  
 DR EMBL; AF043726; AAD21792.1; -.  
 DR MGD; MGI:1338034; Phf2.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001965; PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 SO SEQUENCE 1096 AA; 120798 MW; D12C8DC10938AAF CRC64;

alignment\_scores:  
 Quality: 77.00 Length: 23  
 Ratio: 4.529 Gaps: 0  
 Percent Similarity: 73.913 Percent Identity: 43.478

alignment\_block:  
 US-09-684-016-48411 x Q9WTU0 ..

Align seg 1/1 to: Q9WTU0 from: 1 to: 1096

5 TATTGTAATGTGAGATGCTTACACCGTATGACCTAATGGCGCAATG 54  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 7 TyrcysValCysArgLeuProTyrAspValThrArgPheMetIleGlnCys 23  
 55 CGAGGGCTGCACCTGACTGG 73  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 23 sAspAlaCysLysAspTrp 29

seq\_name: sp\_human:Q9Y6N4

seq\_documentation\_block:  
 ID Q9Y6N4 PRELIMINARY; PRT; 1099 AA.

AC Q9Y6N4;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE PHD-FINGER PROTEIN.  
 GN GRC5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99160472; PubMed=10051327;  
 RA Hasenpusch-Theil K., Chadwick B.P., Theil T., Heath S.K.,  
 RA Wilkison D.G., Frischauf A.M.;  
 RT "PHF2, a novel PHD finger gene located on human chromosome 9q22."  
 RL Mamm. Genome 10:294-298(1999).  
 DR EMBL; AF043725; AAD21791.1; -.  
 DR InterPro: IPR000561; EGF-like.



```

RA Spier E, Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-T., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195(2000).
RT
DR EMBL: AE003655; AAF53607.1; -.
DR FlyBase: FBgn0032635; CG15141.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR003126; ZnF_UBR1.
DR Pfam: PF02207; zf-UBR1; 1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00396; ZnF-UBR1; 1.
SO SEQUENCE 404 AA; 45965 MW; 467FB279F97E995 CRC64;

alignment_scores:
Quality: 76.50 Length: 28
Ratio: 4.250 Gaps: 2
Percent Similarity: 64.286 Percent Identity: 50.000

alignment_block:
US-09-684-016-48411 x Q9YJE3 ..
Align seg 1/1 to: Q9YJE3 from: 1 to: 404

5 TATTGTAATGTGAGATGCGCTTAC..AACCCGAT.....GA 39
|||||
145 TycySLscysLysAsgprOTYrPpASpPpASpATgThvalgluGI 161
|||||
40 CCTAATGCGCGCAATGCAGGCGTCACTGCTG 73
|||||
161 vAlmetleuGlncysLatlelCysgluAsprtp 172

seq_name: sp_human:Q9HAH2

seq_documentation_block:
ID Q9HAH2 PRELIMINARY; PRT; 790 AA.
AC Q9HAH2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CNNA FLJ11634 F1S, CLONE HEMBA1004275, MODERATELY SIMILAR TO HOMO
DE SAPIENS PHD-FINGER PROTEIN (GRC5) MRNA (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC T1SSUB-WHOLE EMBRO, MAINLY HEAD;
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Suano S., Shiratori A., Sudo H.,
RA Nagatsuna M., Hosori T., Kaku Y., Kodaira H., Kondo H., Sugawara
RA Takahashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi
RA Katanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.
RA Yamamoto J., Wakamatsu T.;
RA Ninomiya K., Iwayanagi T.;
RT "NEDD human cDNA sequencing project.";
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK021696; BBI1877.1; -.
DR InterPro: IPR000361; EGF-like.
DR InterPro: IPR001965; PHD.
DR Pfam: PF00628; PHD; 1.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS01186; EGF-2; UNKNOWN_1.
DR NON_TER
FT 790
SQ SEQUENCE 790 AA; 88642 MW; D4D21D2A766C671A CRC64;

```



alignment\_scores:                   Quality: 74.00                   Length: 23  
                                  Ratio: 4.625                   Gaps: 0  
Percent Similarity: 69.565       Percent Identity: 43.478

alignment\_block:  
US-09-684-016-48411 x 09HAH2

Align seg 1/1 to: 09HAH2 from: 1 to: 790

```
5 TATGTGTAATGTGATGACCTTACACCGATGACCTAATGGCGCAATG 54
||||| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
7 Tyricysleucysargleuprototyaspvalthrargphenetileglucy 23
55 CGAGGCTGCACACTGCTGG 73
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
23 SASPMETCysglinsprtp 29
```

seq\_name: sp\_human:09UPP1

seq\_documentation\_block:  
ID 09UPP1           PRELIMINARY;           PRT; 1084 AA.

```
AC 09UPP1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE KIAA1111 PROTEIN (FRAGMENT).
GN KIAA1111.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=93937452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hiroswa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
DR EMBL; AB029034; BAA83063.1; -.
DR Interpro: IPR000561; EGF-like.
DR Interpro: IPR001965; PHD.
DR Pfam: PF00628; PHD; 1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
DR SMART; SM00249; PHD; 1.
FT NON_TER 1
FT SEQUENCE 1084 AA; 120376 MW; D609EBA16BE71A3E CRC64;
```

alignment\_scores:                   Quality: 74.00                   Length: 23  
                                  Ratio: 4.625                   Gaps: 0  
Percent Similarity: 69.565       Percent Identity: 43.478

alignment\_block:  
US-09-684-016-48411 x 09UPP1

Align seg 1/1 to: 09UPP1 from: 1 to: 1084

```
5 TATGTGTAATGTGATGACCTTACACCGATGACCTAATGGCGCAATG 54
||||| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
67 Tyricysleucysargleuprototyaspvalthrargphenetileglucy 83
55 CGAGGCTGCACACTGCTGG 73
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
83 SASPMETCysglinsprtp 89
```

seq\_name: sp\_fungi:Q03012

seq\_documentation\_block:

```
ID 003012           PRELIMINARY;           PRT; 353 AA.
AC 003012;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LP18P.
GN YPL138C OR LP18C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90360986; PubMed=2167832;
RA Tzagoloff A., Capitanio N., Nobrega M.P., Gatti D.;
RT "Cytochrome oxidase assembly in yeast requires the product of COX11, a
RT homolog of the P. denitrificans protein encoded by ORF3."
RL EMBO J. 9:2759-2764(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93140769; PubMed=8423796;
RA Brigati C., Kurtz S., Balderes D., Vidali G., Shore D.;
RT "An essential yeast gene encoding a TTAGGG repeat-binding protein.";
RL Mol. Cell. Biol. 13:1306-1314(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93233668; PubMed=8386320;
RA Irie K., Takase M., Lee K.S., Levin D.E., Araki H., Matsumoto K.,
RA Oshima Y.;
RT "MKR1 and MKR2, which encode Saccharomyces cerevisiae mitogen-
RT activated protein kinase-kinase homologs, function in the pathway
RT mediated by protein kinase C.";
RL Mol. Cell. Biol. 13:3076-3083(1993).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=94287711; PubMed=8017104;
RA Jiang B., Brown J.L., Sheraton J., Fortin N., Bussey H.;
RT "A new family of yeast genes implicated in ergosterol synthesis is
RT related to the human oxysterol binding protein.";
RL Yeast 10:341-353(1994).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=95320205; PubMed=7597071;
RA Shail N., Watkins P.A., Valle D.;
RT "PXAL, a possible Saccharomyces cerevisiae ortholog of the human
RT adrenoleukodystrophy gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:6012-6016(1995).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Anseorge W.,
RA Araujo R., Aparicio A., Barrell B., Bedcock K., Benes V., Botstein D.,
RA Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E.,
RA Churcher C., Coster F., Davis K., Davis R.W., Dietrich F.S.,
RA Delius H., DiPaolo T., Dubois E., Dusterhoft A., Duncan M., Floeth M.,
RA Fortin N., Friesen J.D., Fritz C., Goffeau A., Hall J., Heblung U.,
RA Heumann K., Hilbert H., Hillier L., Hunkeler-Smith S., Hyman R.,
RA Johnston M., Kalman S., Kleine K., Komp C., Kurdi O., Lashkari D.,
RA Lew H., Lin A., Lin D., Louis E.J., Marathe R., Messenguy F.,
RA Mewes H.W., Mitrilipati S., Moestl D., Muller-Auer S., Namath A.,
RA Newnich U., Oefner P., Pearson D., Petel F.X., Pohl T.M.,
RA Purnelle P., Schafer M., Scharfe M., Scherens B., Schramm S.,
RA Schroeder M., Sidon A.M., Tettelin H., Urestrazaru L.A., Ushinsky S.,
RA Vierendeels F., Vissers S., Voss H., Walsh S.V., Wandut R., Wang Y.,
RA Wedler E., Wedler H., Winnett E., Zhong W.W., Zollner A., Vo D.H.,
RA Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:0-0(0).
RN [7]
RP SEQUENCE FROM N.A.
RA Swartzman E., Viswanathan M.N., Emerick A.E., Thorner J.;
RT Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RN [8]
```



RP SEQUENCE FROM N.A.  
 RA Hall J., Depaulo T., Ahmed A., Bussey H., Fortin N., Friesen J.D.,  
 RA Storms R.K., Vo D.H., Wang Y., Minnett E.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bussey H.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RA Jia Y., Cherry J.M.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U43703; AAB68222.1; -.  
 DR SGD: S0006059; tPL138C.  
 DR InterPro: IPR001965; PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 SQ SEQUENCE 353 AA; 41468 MW; F4B8DFE0CB358CA4 CRC64;

alignment\_scores:  
 Quality: 66.50 Length: 23  
 Ratio: 4.156 Gaps: 1  
 Percent Similarity: 69.565 Percent Identity: 47.826

alignment\_block:  
 US-09-684-016-48411 x Q03012 ..

Align seg 1/1 to: Q03012 from: 1 to: 353

5 TATTGTAATGTGAGATGCTTACCAACCGATGACTATGGCGCAATG 54  
 ||||| |||:: ||| ::::| |||||::: ||  
 24 TYRCYSILCYSLYSARPro...AspTYRGLYGLULeuMetValGLYCY 39  
 55 CGAGGGCTGCACCTGACTGG 73  
 |::| ||||| |||||  
 39 SASpGLYCYSAspAspTrp 45

seq\_name: sp\_fung1:074508

seq\_documentation\_block:  
 ID 074508 PRELIMINARY; PRT; 424 AA.  
 AC 074508;

DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE PUTATIVE TRANSCRIPTIONAL REGULATORY PROTEIN, PHD FINGER CONTAINING.  
 GN SPOCC594.05C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_Taxid=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;  
 RA Rieger M., Lyne M., Rajandream M.A., Barrell B.G.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL031523; CAA20664.1; -.  
 DR InterPro: IPR001965; PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 SQ SEQUENCE 424 AA; 48683 MW; 2963605C3DFCC0B9 CRC64;

alignment\_scores:  
 Quality: 62.50 Length: 25  
 Ratio: 4.167 Gaps: 2  
 Percent Similarity: 60.000 Percent Identity: 48.000

alignment\_block:  
 US-09-684-016-48411 x 074508 ..

Align seg 1/1 to: 074508 from: 1 to: 424  
 5 TATTGTAATGTGAGATGCTTACCAACCGATGAC.....CTAATGCC 48  
 ||||| |||:: ||| ::::| |||||::: ||  
 120 TYRCYSILCYSGln.....LysProAspAspGLYSerTrpMetLeu 133  
 49 GCATGCGAGGCTGCACCTGACTGG 73  
 ||::| ||||| |||||  
 133 uGLYCYSAspGLYCYSGLuAspTrp 141

seq\_name: sp\_invertebrate:045410

seq\_documentation\_block:  
 ID 045410 PRELIMINARY; PRT; 405 AA.  
 AC 045410;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE F26H11.3A PROTEIN.  
 GN F26H11.3A.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Barlow K.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berts M.,  
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkes T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans".  
 RL Nature 368:32-38(1994).  
 DR EMBL: Z81515; CAB04198.1; -.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001487; Bromodomain.  
 DR InterPro: IPR001965; PHD.  
 DR Pfam: PF00439; Bromodomain; 1.  
 DR PRINTS: PR00503; BROMODOMAIN.  
 DR PROSITE: PS50014; BROMODOMAIN\_2; 1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR SMART: SM00297; BROMO; 1.  
 DR SMART: SM00249; PHD; 2.  
 SQ SEQUENCE 405 AA; 46199 MW; DE6BE7384A5AE82F CRC64;

alignment\_scores:  
 Quality: 62.00 Length: 23  
 Ratio: 4.429 Gaps: 0  
 Percent Similarity: 60.870 Percent Identity: 34.783

alignment\_block:  
 US-09-684-016-48411 x 045410 ..

Align seg 1/1 to: 045410 from: 1 to: 405

5 TATTGTAATGTGAGATGCTTACCAACCGATGACTGATGCGCAATG 54  
 ||||| |||:: |||||::: ||  
 172 TYRCYSValCYSGlnLysProTrpAspAspTrpLysPheTYValGLYCY 188  
 55 CGAGGGCTGCACCTGACTGG 73



188 saspsercysglncltytrp 194

seq\_name: sp.invertebrate:Q9NMC1

seq\_documentation\_block:  
ID Q9NMC1 PRELIMINARY; PRT; 451 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE F26H11.3C PROTEIN.

GN F26H11.3C.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_Taxid=6239;

RA Barlow K.;

RP SEQUENCE FROM N.A.

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN (2)

RP SEQUENCE FROM N.A.

RX MEDLINE=9069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for

investigating biology."

RL Science 282:2012-2018(1998).

DR EMBL; 281515; CAB54234.2; -

DR InterPro: IPR001487; Bromodomain.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001965; PHD.

DR Pfam: PF00439; bromodomain; 1.

DR Pfam: PF00628; PHD; 2.

DR PRINTS: PR00503; BROMODOMAIN.

DR SMART: SM00297; BROMO; 1.

DR SMART: SM00249; PHD; 2.

DR PROSITE: PS50014; BROMODOMAIN\_2; 1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

OX NCBI\_Taxid=6239;

RA Barlow K.;

RP SEQUENCE FROM N.A.

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN (2)

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Altschough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,

RA Smailon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierley-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans."

RL Nature 368:32-38(1994).

DR EMBL; 281515; CAB04195.1; -

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001487; Bromodomain.

DR InterPro: IPR001965; PHD.

DR Pfam: PF00439; bromodomain; 1.

DR Pfam: PF00628; PHD; 2.

DR PRINTS: PR00503; BROMODOMAIN.

DR PROSITE: PS50014; BROMODOMAIN\_2; 1.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.

DR SMART: SM00297; BROMO; 1.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

DR SMART: SM00249; PHD; 2.

alignment\_scores:

Quality: 62.00

Ratio: 4.429

Percent Similarity: 60.870

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

Percent Identity: 34.783

alignment\_block:

US-09-684-016-48411 x Q9NMC1

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

alignment\_block:

US-09-684-016-48411 x Q9NMC1

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451

Align seg 1/1 to: Q9NMC1 from: 1 to: 451



DR EMBL: AB032251; BAA89208.1; -  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001487; Bromodomain.  
 DR InterPro: IPR001965; PHD.  
 DR Pfam: PF00439; bromodomain; 1.  
 DR Pfam: PF00628; PHD; 2.  
 DR PRINTS: PR00503; BROMODOMAIN.  
 DR PROSITE: PS00633; BROMODOMAIN\_1; 1.  
 DR PROSITE: PS0014; BROMODOMAIN\_2; 1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR SMART: SM00297; BROMO; 1.  
 DR SMART: SM00249; PHD; 2.  
 SQ SEQUENCE 2781 AA; 311210 MW; 7B3B0C1C4C3CFBE CRC64;

alignment\_scores:  
 Quality: 60.00 Length: 23  
 Ratio: 4.286 Gaps: 0  
 Percent Similarity: 60.870 Percent Identity: 34.783

alignment\_block:  
 US-09-684-016-48411 x Q9UIG2 ..

Align seg 1/1 to: Q9UIG2 from: 1 to: 2781

5 TATGTAAATGAGATGCGCTTACAAACCTGATGACCTAATGGCGCAATG 54  
 ||||| |||:: |||||::: |||::: |||  
 2604 TYRCYSILecyslysthrProtyrAspGluSerlystherylleGlyCy 2620  
 55 CGAGGCGCTGCACCTGACTG 73  
 |::: |||::: |||  
 2620 SASPARCYSGlnAsnTrp 2626  
 seq\_name: sp\_human:014687

seq\_documentation\_block:  
 ID 014687; PRELIMINARY; PRT; 4957 AA.  
 AC 014687;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ALR.  
 GN ALR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97388474; PubMed=9247308;  
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
 RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
 RA Canaan E.;  
 RT "Structure and expression pattern of human ALR, a novel gene with  
 RT strong homology to ALL-1 involved in acute leukemia and to Drosophila  
 RT trithorax.";  
 RL Oncogene 15:549-560(1997).  
 DR EMBL: AF010404; AAC51735.1; -  
 DR InterPro: IPR000910; HMG\_12\_box.  
 DR InterPro: IPR001214; SET.  
 DR InterPro: IPR001822; Recombinase.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR003616; PostSET.  
 DR InterPro: IPR003888; FYRICH\_N.  
 DR InterPro: IPR003889; FYRICH\_C.  
 DR Pfam: PF00628; PHD; 3.  
 DR Pfam: PF00856; SET; 1.  
 DR PRINTS: PR01217; PRICHEXTENSN.  
 DR PROSITE: PS00398; RECOMBINASES\_2; UNKNOWN\_1.  
 DR PROSITE: PS50280; SET; 2.  
 DR SMART: SM00542; FYRC; 1.  
 DR SMART: SM00541; FYRN; 1.  
 SQ SEQUENCE 5262 AA; 564181 MW; 26B7C74CAD417E44 CRC64;

DR SMART: SM00398; HMG; 1.  
 DR SMART: SM00249; PHD; 4.  
 DR SMART: SM00508; PostSET; 1.  
 DR SMART: SM00317; SET; 1.  
 SQ SEQUENCE 4957 AA; 531848 MW; 1026562E1419CE8D CRC64;

alignment\_scores:  
 Quality: 60.00 Length: 20  
 Ratio: 4.615 Gaps: 0  
 Percent Similarity: 65.000 Percent Identity: 45.000

alignment\_block:  
 US-09-684-016-48411 x 014687 ..

Align seg 1/1 to: 014687 from: 1 to: 4957

14 TGTGAGATGCGCTTACAAACCTGATGACCTAATGGCGCAATGGAGGCGTG 63  
 |||::: |||||::: |||||::: |||||::: ||  
 930 CysHisAlaProtyrValGIuGlnAspLeuLeuIleGlnCysArgHisCy 946  
 64 CACTGACTG 73  
 |||  
 946 SGIuArgTrp 949  
 seq\_name: sp\_human:014686

seq\_documentation\_block:  
 ID 014686; PRELIMINARY; PRT; 5262 AA.  
 AC 014686;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ALR.  
 GN ALR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97388474; PubMed=9247308;  
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
 RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
 RA Canaan E.;  
 RT "Structure and expression pattern of human ALR, a novel gene with  
 RT strong homology to ALL-1 involved in acute leukemia and to Drosophila  
 RT trithorax.";  
 RL Oncogene 15:549-560(1997).  
 DR EMBL: AF010403; AAC51734.1; -  
 DR InterPro: IPR000910; HMG\_12\_box.  
 DR InterPro: IPR001214; SET.  
 DR InterPro: IPR001822; Recombinase.  
 DR InterPro: IPR001841; Znf\_ring.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR003616; PostSET.  
 DR InterPro: IPR003888; FYRICH\_N.  
 DR InterPro: IPR003889; FYRICH\_C.  
 DR Pfam: PF00628; PHD; 5.  
 DR Pfam: PF00856; SET; 1.  
 DR PRINTS: PR01217; PRICHEXTENSN.  
 DR PROSITE: PS00398; RECOMBINASES\_2; UNKNOWN\_1.  
 DR PROSITE: PS50280; SET; 2.  
 DR SMART: SM00542; FYRC; 1.  
 DR SMART: SM00541; FYRN; 1.  
 DR SMART: SM00398; HMG; 1.  
 DR SMART: SM00249; PHD; 7.  
 DR SMART: SM00508; PostSET; 1.  
 DR SMART: SM00184; RING; 3.  
 DR SMART: SM00317; SET; 1.  
 SQ SEQUENCE 5262 AA; 564181 MW; 26B7C74CAD417E44 CRC64;



## alignment\_scores:

Quality: 60.00 Length: 20  
Ratio: 4.615 Gaps: 0  
Percent Similarity: 65.000 Percent Identity: 45.000

## alignment\_block:

US-09-684-016-48411 x 014686 ..

Align seg 1/1 to: 014686 from: 1 to: 5262

14 TGTGAGATGCTTACCAACCTGATGCGCAATGCGAGGCTG 63  
||||| ||||| ::||| ||||| ||||| ||||| |||||  
1235 CysHsAlaProTyValGluGluAspLeuIleGlnCysArgHnScy 1251  
64 CACTGACTGG 73  
|||  
1251 sgluArgTrrp 1254  
seq\_name: sp\_invertebrate:Q22662

## seq\_documentation\_block:

ID Q22662 PRELIMINARY; PRT; 416 AA.  
AC Q22662;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE T22C1.1 PROTEIN.  
GN T22C1.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McMurtry A.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Ritken L., Koopra A., Saunders D., Showkhen R.,  
RA Smalton N., Smith A., Sonhammer E., Staden K., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans";  
RL Nature 368:32-38(1994).  
DR EMBL: Z75550; CAA9920.1; -.  
DR InterPro: IPR003126; ZNF\_UBR1.  
DR Pfam: PF02207; zf-UBR1.1.  
DR SMART: SM00396; ZNF\_UBR1.1.  
SO SEQUENCE 416 AA; 47619 MW; 3950A83C4D77E9D2 CRC64;

## alignment\_scores:

Quality: 59.50 Length: 27  
Ratio: 3.719 Gaps: 1  
Percent Similarity: 59.259 Percent Identity: 37.037

## alignment\_block:

US-09-684-016-48411 x 022662 ..

Align seg 1/1 to: Q22662 from: 1 to: 416

2 AGGTATGTAATGTGAGATGCTTACACCCGTGATGAC.....CT 42  
:||||| ||||| ||| :|||  
197 LysPheCysThrCysAspValPheTyrProAspGluAspGlyGlySGI 213

43 AATGGCGCATGCGAGGCTGCACACTGG 73

||||| ||||| ||| |||||  
213 uLeuGlnCysGluIleCysGluAspTrrp 223

seq\_name: sp\_plant:Q9T0A1

## seq\_documentation\_block:

ID Q9T0A1 PRELIMINARY; PRT; 443 AA.  
AC Q9T0A1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE HYPOHETICAL 49.7 KDA PROTEIN.  
GN T32A16.30 OR AT4G23860.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Zimmermann W., Gruenisen A., Wambutt R., Bancroft I.,  
RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.,  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Zimmermann W., Gruenisen A., Wambutt R., Kalicki J., Wohlmann P.,  
RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.,  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL078468; CAB43886.1; -.  
DR EMBL: AL161560; CAB81304.1; -.  
DR InterPro: IPR001965; PHD.  
DR InterPro: IPR003126; ZNF\_UBR1.  
DR Pfam: PF00628; PHD.1.  
DR Pfam: PF02207; zf-UBR1.1.  
DR SMART: SM00396; PHD.1.  
DR SMART: SM00396; ZNF\_UBR1.1.  
KW Hypothetical protein.  
SO SEQUENCE 443 AA; 49704 MW; 7DA998CA6ED2904D CRC64;

## alignment\_scores:

Quality: 59.50 Length: 29  
Ratio: 4.250 Gaps: 2  
Percent Similarity: 48.276 Percent Identity: 44.828

## alignment\_block:

US-09-684-016-48411 x Q9T0A1 ..

Align seg 1/1 to: Q9T0A1 from: 1 to: 443

5 TATGTAAATGTGAGATGCTTACACCCGTGATGAC..... 40  
||||| ||||| ||||| |||||  
128 TyrcysthrCysAspArgProTyr...ProAspProAsnValGluGluGI 143  
41 .....CTAATGGCGCATGCGAGGCTGCACACTGG 73  
||| ||||| ||| |||||  
143 nValGluMetIleGlnCysCysLeuGluAspTrrp 155

seq\_name: sp\_invertebrate:Q9YRV6

## seq\_documentation\_block:

ID Q9YRV6 PRELIMINARY; PRT; 662 AA.  
AC Q9YRV6;



alignment_scores:	Quality: 57.50	Length: 33
	Ratio: 2.212	Gaps: 1
	Percent Similarity: 78.788	Percent Identity: 36.364
alignment_block:		
US-09-684-016-48411/rev x GQ9RV6		.

```

seq_name: sp_fungi:Q9UT79
seq_documentation_block:
ID Q9UT79 PRELIMINARY; PRT: 1588 AA.
Q9UT79;
01-MAY-2000 (TREMBLrel. 13, Created)
01-MAY-2000 (TREMBLrel. 13, Last sequence update)
01-JUN-2001 (TREMBLrel. 17, Last annotation update)
PUTATIVE TRANSCRIPTIONAL REGULATORY PROTEIN, PHD FINGER.
SPAC343.11C.
Schizosaccharomyces pombe (Fission yeast).
Eukaryota; fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
NCBI_TaxID=4896;
[1]
SEQUENCE FROM N.A.A.
STRAIN-972H:
Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
EMBL: AL109739; CAB52274.1; -.
InterPro: IPR0000345; Cytochrome_bind.
InterPro: IPR0000561; EGF-like.
InterPro: IPR001365; PHD.
InterPro: IPR003347; UMG_C.
InterPro: IPR003349; UMG_N.
Pfam: PF00628; PHD; 3.
Pfam: PF02373; jmjC.1.
PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
PROSITE: PS0186; EGF_2; UNKNOWN_1.
SMART: SM00545; jmjN; 1.
SMART: SM00249; PHD; 3.
SEQUENCE 1588 AA; 180350 MW; 2D18248ABB3DEC55 CRC64;

```

```
seq_name: sg_human:Q9BOW0
seq_documentation_block:
  Q9BOW0 PRELIMINARY; PRT; 379 AA.
  Q9BOW0:
    01-JUN-2001 (TREMBLrel. 17, Created)
    01-JUN-2001 (TREMBLrel. 17, Last sequence update)
    01-JUN-2001 (TREMBLrel. 17, Last annotation update)
    DJ88517.9.3 (DEATH ASSOCIATED TRANSCRIPTION FACTOR 1 (CONRAINS))
```



```
DE KIAA0333), ISOFORM 3) (FRAGMENT).
GN DAFI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035669; CAC28883.1; -.
FT NON_TER
SQ SEQUENCE 379 AA; 41245 MW; 3B14597F9243C6D5 CRC64;

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 3.531       Gaps: 1
  Percent Similarity: 69.565      Percent Identity: 39.130

alignment_block:
US-09-684-016-48411 x Q9BQW0  ..

Align seg 1/1 to: Q9BQW0 from: 1 to: 379

5 TATGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCGCAATG 54
||||| |||:|||||:||||| |||: |||
87 TyrcysllecySarGlnProHisAsn...AsnarGpneMetIleCyCy 102
55 CGAGGCTGCACACTGCTGG 73
|:: ||| ::|||
102 sasparGcysgluGluTrp 108

seq_name: sp_human:Q9BQW0

seq_documentation_block:
ID Q9BQW0 PRELIMINARY; PRT; 544 AA.
AC Q9BQW0;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE HYPOTHETICAL 59.5 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC000770; AAH00770.1; -.
KW Hypothetical protein.
SQ SEQUENCE 544 AA; 59487 MW; 193341A7ED5FDA CRC64;

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 3.531       Gaps: 1
  Percent Similarity: 69.565      Percent Identity: 39.130

alignment_block:
US-09-684-016-48411 x Q9BQW0  ..

Align seg 1/1 to: Q9BQW0 from: 1 to: 544

5 TATGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCGCAATG 54
||||| |||:|||||:||||| |||: |||
270 TyrcysllecySarGlnProHisAsn...AsnarGpneMetIleCyCy 285
55 CGAGGCTGCACACTGCTGG 73
|:: ||| ::|||
285 sasparGcysgluGluTrp 291

seq_name: sp_human:Q9BQW0
```

```
seq_name: sp_human:Q9NUM8

seq_documentation_block:
ID Q9NUM8 PRELIMINARY; PRT; 562 AA.
AC Q9NUM8;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE CDNA FL111265 FIS, CLONE PLACE1009158.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK002127; BAA92094.1; -.
DR InterPro; IPR001965; PHD.
DR Pfam; PF0628; PHD; 1.
DR SMART; SM00249; PHD; 1.
SQ SEQUENCE 562 AA; 61425 MW; 3BA79F8DEF9D3D9C CRC64;

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 3.531       Gaps: 1
  Percent Similarity: 69.565      Percent Identity: 39.130

alignment_block:
US-09-684-016-48411 x Q9NUM8  ..

Align seg 1/1 to: Q9NUM8 from: 1 to: 562

5 TATGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCGCAATG 54
||||| |||:|||||:||||| |||: |||
270 TyrcysllecySarGlnProHisAsn...AsnarGpneMetIleCyCy 285
55 CGAGGCTGCACACTGCTGG 73
|:: ||| ::|||
285 sasparGcysgluGluTrp 291

seq_name: sp_human:Q9BRC0

seq_documentation_block:
ID Q9BRC0 PRELIMINARY; PRT; 562 AA.
AC Q9BRC0;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE HYPOTHETICAL 61.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-COLON ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004237; AAH04237.1; -.
KW Hypothetical protein.
SQ SEQUENCE 562 AA; 61353 MW; 91A71468F9C5C12 CRC64;
```



alignment\_scores:  
 Quality: 56.50 Length: 23  
 Ratio: 3.531 Gaps: 1  
 Percent Similarity: 69.565 Percent Identity: 39.130

alignment\_block:

US-09-684-016-48411 x Q9W00

Align seg 1/1 to: Q9W00 from: 1 to: 562

5 TATTGTAATGATGAGATGCTTACACCCGATGACCTATGCGCATG 54  
 ||||| |||:|||||:||||| :||| :|||  
 270 TYRCYSILecysarglnProHisasn...AsnargPheMetIleCysCy 285  
 55 CGAGGCGTGCACCTGACTGG 73  
 |||: ||| :|||  
 285 sasparGcysgluGluTirp 291

seq\_name: sp\_todent:Q9W00

seq\_documentation\_block:

ID Q9W00 PRELIMINARY; PRT; 614 AA.  
 AC Q9W00:  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE DEATH INDUCER-OBLITERATOR-1 (FRAGMENT).  
 GN DIDOI OR DIO-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99324176; PubMed=10393935;  
 RA Garcia-Domingo D., Leonardo E., Grandien A., Martinez P., Albar J.P.,  
 RA Izpisua-Belmonte J.C., Martinez-A.C.  
 RT "DIO-1 is a gene involved in onset of apoptosis in vitro, whose  
 RT misexpression disrupts limb development."  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:7992-7997(1999).  
 DR EMBL: AJ238332; CAB48401.1; -  
 DR MGD: MGI:1344352; Didoi1.  
 DR InterPro: IPR001965; PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 FT NON\_TER 614 614  
 SEQUENCE 614 AA; 67366 MW; F3E6B8970333AADA CRC64;

alignment\_scores:  
 Quality: 56.50 Length: 23  
 Ratio: 3.531 Gaps: 1  
 Percent Similarity: 69.565 Percent Identity: 39.130

alignment\_block:

US-09-684-016-48411 x Q9W00

Align seg 1/1 to: Q9W00 from: 1 to: 614

5 TATTGTAATGATGAGATGCTTACACCCGATGACCTATGCGCATG 54  
 ||||| |||:|||||:||||| :||| :|||  
 267 TYRCYSILecysarglnProHisasn...AsnargPheMetIleCysCy 282  
 55 CGAGGCGTGCACCTGACTGG 73  
 |||: ||| :|||  
 282 sasparGcysgluGluTirp 288

seq\_name: sp\_vertebrate:Q9W664

seq\_documentation\_block:

ID Q9W664 PRELIMINARY; PRT; 641 AA.  
 AC Q9W664;

DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE P73.  
 OS Barbus barbus (Barbel).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Cyprininae; Barbus.  
 OX NCBI\_TaxID=40830;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bhaskaran A., May D., Rand-Weaver M., Tyler C.R.;  
 RT "Fish p73, ancestral p53?,"  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF043641; AAD27752.1; -  
 DR HSSP: O15350; ICOK.  
 DR InterPro: IPR002117; P53.  
 DR InterPro: IPR001660; SAM.  
 DR Pfam: PF00870; P53; 1.  
 DR PRINTS: PR00386; P53SUPPRESSR.  
 DR ProDom: PD002681; P53; 1.  
 DR SMART: SM00454; SAM; 1.  
 SEQUENCE 641 AA; 71090 MW; 4A215B21AEB16EAD CRC64;

alignment\_scores:  
 Quality: 56.50 Length: 45  
 Ratio: 2.260 Gaps: 2  
 Percent Similarity: 55.556 Percent Identity: 31.111

alignment\_block:

US-09-684-016-48411 x Q9W664

Align seg 1/1 to: Q9W664 from: 1 to: 641

5 TATTGTAATGATGAGATGCTTACACCCGATGACCTATGCGCAA.. 52  
 :||| ||||| ||||| :||| :|||  
 487 HiscysThrProProProtyrAsnProAspProSerLeuValSerph 503  
 53 .....TGCAGGCGTGCACCTGACCTGTAAGTTTCACATG 86  
 |||:|||||:||||| :||| :|||  
 503 eleuthinSerLeuGlyCysGlnAsnGlySerPyrPheThrSerGln. 519  
 87 TAAGGTGATGATCTTATCTTCATTTATCCACACAA 121  
 |||:|||||:||||| :||| :|||  
 520 .....GlyLeuGlnSerValTyrHisLeuGln 528

seq\_name: sp\_human:O15043

seq\_documentation\_block:

ID O15043 PRELIMINARY; PRT; 991 AA.  
 AC O15043;  
 DT 01-JAN-1998 (TReMBLrel. 05, Created)  
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE KIAA0333 (FRAGMENT).  
 GN KIAA0333.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=BRIN;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro."  
 RL DNA Res. 4:141-150(1997).  
 DR EMBL: AB002331; BAA20791.1; -  
 DR InterPro: IPR001052; Rubredoxin.



DR InterPro: IPR001209; Ribosomal\_S14.  
DR InterPro: IPR001965; PHD.  
DR InterPro: IPR003618; TFS2\_cent.  
DR Pfam: PF00628; PHD; 1.  
DR PRINTS: PR00163; RUBREDOXIN.  
DR PROSITE: PS00527; RIBOSOMAL\_S14; UNKNOWN\_1.  
DR SMART: SM00249; PHD; 1.  
DR SMART: SM00510; TFS2M; 1.  
FT NON\_TER 1  
SQ SEQUENCE 991 AA; 107368 MW; 3483D97D3290A7A1 CRC64;

## alignment\_scores:

Quality: 56.50 Length: 23  
Ratio: 3.531 Gaps: 1  
Percent Similarity: 69.565 Percent Identity: 39.130

## alignment\_block:

US-09-684-016-48411 x 015043 ..

Align seg 1/1 to: 015043 from: 1 to: 991

5 TATTGTAATGTGAGATGCCCTTACACCGCTGATGACCTAATGCGCAATG 54  
||||| |||:|||||:||||| :| :|:|||||  
36 TyrcysIleCysArgGlnProHisAsn...AsnArgPheMetIleCysCy 51  
55 CGAGGGCTGCACCTGACTGCTG 73  
|:|:| | | :|:|:| | |  
51 SASPARGCGSGlUGlUTP 57

seq\_name: sp\_human:Q9H4G6

## seq\_documentation\_block:

ID Q9H4G6 PRELIMINARY; PRT; 1042 AA.  
AC Q9H4G6;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
DE DJ88517.9.1 (NOVEL PROTEIN SIMILAR TO MOUSE DEATH INDUCER OBLITERATOR  
DE 1 (DIO-1) (CONTAINS KIAA0333) (ISOFORM 1)) (FRAGMENT).  
GN DJ88517.9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Smith M.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL035669; CAC12756.1; -  
DR InterPro: IPR001965; PHD.  
DR InterPro: IPR001209; Ribosomal\_S14.  
DR InterPro: IPR003618; TFS2\_cent.  
DR Pfam: PF00628; PHD; 1.  
DR SMART: SM00249; PHD; 1.  
DR SMART: SM00510; TFS2M; 1.  
DR PROSITE: PS00527; RIBOSOMAL\_S14; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 1042 AA; 112916 MW; F1ADB40059F09742 CRC64;

## alignment\_scores:

Quality: 56.50 Length: 23  
Ratio: 3.531 Gaps: 1  
Percent Similarity: 69.565 Percent Identity: 39.130

## alignment\_block:

US-09-684-016-48411 x Q9H4G6 ..

Align seg 1/1 to: Q9H4G6 from: 1 to: 1042

5 TATTGTAATGTGAGATGCCCTTACACCGCTGATGACCTAATGCGCAATG 54  
||||| |||:|||||:||||| :| :|:|||||  
||| :|:|:| :|:|:| :|:|:| :|

87 TyrcysIleCysArgGlnProHisAsn...AsnArgPheMetIleCysCy 102  
55 CGAGGGCTGCACCTGACTGCTG 73  
|:|:| | | :|:|:| | |  
102 SASPARGCGSGlUGlUTP 108

seq\_name: sp\_archaea:Q9YEG2

## seq\_documentation\_block:

ID Q9YEG2 PRELIMINARY; PRT; 544 AA.  
AC Q9YEG2;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
DE HYPOPHETICAL 55.8 KDA PROTEIN APE0614.  
GN APE0614.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;  
OC Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,  
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL: AP000060; BAA79584.1; -  
DR InterPro: IPR000566; Lipocln\_cyFABP.  
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 544 AA; 55839 MW; 14028D4D558A7891 CRC64;

## alignment\_scores:

Quality: 56.00 Length: 26  
Ratio: 3.111 Gaps: 0  
Percent Similarity: 69.231 Percent Identity: 38.462

## alignment\_block:

US-09-684-016-48411/rev x Q9YEG2 ..

Align seg 1/1 to: Q9YEG2 from: 1 to: 544

116 TGGATTAATTTGAAGATTAAGTATACCTTACATGTGAACCTTACCAGTCA 67  
||||| :|:| :|:| :|:|:|:|:|:| :| :|:|:|:|:|  
179 TrpIupTroIleaspAlaValPheIysValGluIleGluLeuAspArgSe 195  
66 GTGCAGCCCTGCGATTCGCGCCATTAGG 39  
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
195 CySserProArgIleAlaIalaLeuGly 204

seq\_name: sp\_bacteria:Q9RLR4

## seq\_documentation\_block:

ID Q9RLR4 PRELIMINARY; PRT; 363 AA.  
AC Q9RLR4;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)  
DE LVH310 PROTEIN.  
GN LVH310.  
OS Legionella pneumophila.  
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;  
OC Legionellaceae; Legionella.  
OX NCBI\_TaxID=446;  
RN [1]







```

RX MEDLINE-98069011; PubMed=9405337;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL: AB06699; BAB1682.1; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001965; PHD.
DR Pfam: PF00628; PHD; 2.
DR SMART: SM00249; PHD; 3.
DR PROSITE: PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN 1.
SO SEQUENCE 1516 AA; 170012 MW; FFAA8593BA93CEA CAC64;

alignment_scores:
Quality: 55.00 Length: 20
Ratio: 4.583 Gaps: 0
Percent Similarity: 60.000 Percent Identity: 40.000

alignment_block:
US-09-684-016-48411 x Q9FN15 ..
..

Align seg 1/1 to: Q9FN15 from: 1 to: 1516

14 TGTGAGATGCGCTTACACCGTATGACCTAATGCGCATGCGAGGCGTG 63
|||||.....:|||||.....:|||||
1198 CyslystLeupProTyrAspProGlyLeuThrTyrTleHisCysThrSerC 1214
64 CACTGACTGG 73
|.....|
1214 saspmetrrp 1217

seq_name: sp.invertebrate:Q9U263

seq_documentation_block:
ID Q9U263 PRELIMINARY; PRT; 1645 AA.
AC Q9U263;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Y1HA4.12 PROTEIN.
GN Y1HA4.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae.
OC Rhabditidae; Pelodermae; Caenorhabditis.
OX NCBI_TaxID=6239;
RX [1]
RN SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RL EMBL: AL132952; CAB63382.1; -.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001965; PHD.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF00856; SET; 1.
DR PROSITE: PS00280; SET; 1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00317; SET; 1.
SO SEQUENCE 1645 AA; 181664 MW; 74BF8830788BBD57 CAC64;

alignment_scores:
Quality: 54.50 Length: 22
Ratio: 4.192 Gaps: 1

```



Percent Similarity: 59.091 Percent Identity: 40.909

## alignment\_block:

US-09-684-016-48411 x 09U263

Align seg 1/1 to: 09U263 from: 1 to: 1645

```

      8 TGTAAATGTGAGATGCTTACACACCTGATGACCTTAATGCGCAATGCGA 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      797 CysHiscysglYmetAspHisglYAspThrIle...GlucYsgl 812
      58 GGGCTGCACACTGACTGG 73
      ||||| |||
      812 uGlyCysLysThrTrp 817

```

seq\_name: sp\_invertebrate:044498

## seq\_documentation\_block:

ID 044498 PRELIMINARY; PRT: 1655 AA.

AC 044498:

DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE F15E6.1 PROTEIN.

CN F15E6.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Altschough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lichten J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,

RA Smaildon J., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RN [2]

SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Miller N., Steillyes L., Bradshaw H., Keppler D.;

RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF038614; AAB92063.1; -

DR InterPro: IPR001214; SET.

DR InterPro: IPR001965; PHD.

DR Pfam: PF00628; PHD; 1.

DR Pfam: PF00856; SET; 1.

DR PROSITE: PS50280; SET; 1.

DR SMART: SM00249; PHD; 1.

DR SMART: SM00317; SET; 1.

SO SEQUENCE 1655 AA; 182580 MW; 0AB49D12CB1C0686 CRC64;

## alignment\_scores:

Quality: 54.50 Length: 22

Ratio: 4.192 Gaps: 1

Percent Similarity: 59.091 Percent Identity: 40.909

## alignment\_block:

US-09-684-016-48411 x 044498

Align seg 1/1 to: 044498 from: 1 to: 1655

```

      8 TGTAAATGTGAGATGCTTACACACCTGATGACCTTAATGCGCAATGCGA 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      821 CysHiscysglYmetAspHisglYAspThrIle...GlucYsgl 836
      58 GGGCTGCACACTGACTGG 73
      ||||| |||
      836 uGlyCysLysThrTrp 841

```

seq\_name: sp\_invertebrate:Q9VR08

## seq\_documentation\_block:

ID Q9VR08 PRELIMINARY; PRT: 3680 AA.

AC Q9VR08:

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE CG15637 PROTEIN.

DN CG15637.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amandlis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brannon R.C., Rogers Y.H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Adabayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kalp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheel F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Massarman D.A., Wellstock G.M., Wellsteinbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RL Science 287:2185-2195(2000).

DR EMBL: AE003576; AAF51000.1; -

DR HSSP: P35555; 1EMN

DR Flybase: FBgn000488; dp

DR InterPro: IPR000152; Asx\_hydroxyl.



```

RA Hostin D., Harvey D., Helman T.J., Hernandez J.R., Houck C.,
RA Dattil M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shue B.C., Sidenk-Ramos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenhach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yen K.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA Zibib R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster." ;
RL Science 287:2185-2195(2000).
DR EMBL: AEO03806; AAF57974.1; -
DR FlyBase; FBgn0034145; CG5065.
SQ SEQUENCE 625 AA; 69682 MW; 8C12CD780F96D5E0 CRC64;

alignment_scores:
    Quality: 54.00      Length: 25
    Ratio: 3.857      Gaps: 1
    Percent Similarity: 56.000      Percent Identity: 36.000

alignment_block:
US-09-684-016-48411 x Q9V7S1 ..

Align seg 1/1 to: Q9V7S1 from: 1 to: 625

5 TATTGTAATGTGAGATG.....CC 24
|||||:|||||:
270 TycrCysAaNCysAaSPaRGThrApsValSerGluValIleTyrAlaProPr 286
25 TTACACACCCGTGATGCACCAATAGGG 49
|||||:|||||:
286 cTyrAnPrnAaSPaSPaSPleIleIeSer 294

seq_name: sp.invertebrate:Q24459

seq_documentation_block:
ID Q24459 PRELIMINARY; PRT; 857 AA.
AC Q24459;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-1996 (TrEMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
DE POLYCOMBLIKE NUCLEAR PROTEIN.
GN PCL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOCENTRIC CHROMOSOME 2, DP CN BW;
RC MEDLINE=95044940; PubMed=7936837;
RA Lonie A., Dandrea R., Pato R., Saint R.;
RT "Molecular characterisation of the Polycomblike gene of Drosophila
RT melanogaster, a trans-acting negative regulator of homeotic gene
RT expression." ;
RL development 120:2629-2636(1994).
CC - SIMILARITY: CONTAINS 2 PHD-FINGER DOMAINS.
DR EMBL: L35133; AAA64457.1; -
DR FlyBase; FBgn0003044; PCL.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR002999; Tudor.

```







DR InterPro: IPR001245; Tyr\_kin.  
 DR Pfam: PF00560; LRR; 2.  
 DR Pfam: PF00699; pkinase; 1.  
 DR PRINTS: PRO0109; LEURICHRPT.  
 DR PRINTS: PRO0109; TYRKINASE.  
 DR SMART: SM00370; LRR; 4.  
 DR SMART: SM00221; STYK; 1.  
 DR SMART: SM00220; S\_TK; 1.  
 DR SMART: SM00219; TYRK; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR ATP-binding: Kinase; Transferase.  
 KW SEQUENCE 570 AA: 63537 MW: 972414BDE3A8DBC0 CRC64;

alignment\_scores:  
 Quality: 53.50 Length: 40  
 Ratio: 2.140 Gaps: 3  
 Percent Similarity: 62.500 Percent Identity: 40.000

alignment\_block:  
 US-09-684-016-48411/rev x Q9VG65 ..

Align seg 1/1 to: Q9VG65 from: 1 to: 570

116 TGGATAAATGAGATGATGACCTTACAT.....GT 82  
 |||||: :|||||:|||||:|||||: ||  
 353 TTPSNAAG...ArgIleSerIleAlaIleuGlyAlaAlaGlyLeuVA 368  
 ||| :|||:|||||:|||||: |||||:  
 81 GAAACTTACACGATGCGACCGCTCGCATTCGCCCATTAAGTCATCAGG 32  
 ||| :|||:|||||:|||||: |||||:  
 368 ILYLEHISGLINCISASNPOLYSILE.....HSAVGA 382  
 |||||:|||||:|||||:|||||: |||||:  
 31 GTGTAGGCACTTCACATT 12  
 |||||:|||||:|||||:|||||: |||||:  
 382 SPVALYSAALAAASNLLE 388

seq\_name: sp\_invertebrate:Q9VG78

seq\_documentation\_block:  
 ID Q9VG78 PRELIMINARY; PRT; 2016 AA.

AC Q9VG78;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE CG6525 PROTEIN.  
 GN CG6525.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brostein P., Brottier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,  
 RA Jalali M., Kalush E., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003695; AAP54807.1;  
 DR FlyBase: FBgn0038041; CG6525.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR003618; TFS2\_cent.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 DR SMART: SM00510; TFS2M; 1.  
 DR SEQUENCE 2016 AA: 221998 MW: FB61A332EA710F5E CRC64;

alignment\_scores:  
 Quality: 53.50 Length: 23  
 Ratio: 3.344 Gaps: 1  
 Percent Similarity: 69.365 Percent Identity: 39.130

alignment\_block:  
 US-09-684-016-48411 x Q9VG78 ..

Align seg 1/1 to: Q9VG78 from: 1 to: 2016

5 TATTGTAAATGAGATGATGACCTTACACCGCTGATGACCGCATG 54  
 :|||:|||||:|||||:|||||: :|||: ||  
 912 TTPSYSLCYSGARGLNPHISASN...ASNAARGPMETILCYCY 927  
 ||| :|||:|||||:|||||: |||||:  
 55 CGAGGCTGCACCTGACTGG 73  
 ||| :|||:|||||:|||||: |||||:  
 927 SASPLEUCYSGLSLSPTR 933

seq\_name: sp\_virus:Q98VX5

seq\_documentation\_block:  
 ID Q98VX5 PRELIMINARY; PRT; 198 AA.

AC Q98VX5;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE ENVELOPE PROTEIN (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_Taxid=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IND. CAL1058;  
 RA Mandal D., Chakrabarti S.;  
 RT "Divergent HIV-1 C subtypes isolated from Calcutta, India.";  
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: A3311641; CAC34551.1; -;  
 FT NON\_TER 1  
 FT NON\_TER 198  
 FT SEQUENCE 198 AA: 22154 MW: 51583AFE0BF1EDC1 CRC64;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:32:46 ; Search time 1177.44 Seconds

(without alignments)  
1122.546 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 caggtatgttaaatgtaga.....tcaattatccacacaca 123

Scoring table: IDENTITY\_NUC

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estom:\*  
5: em\_estopl:\*  
6: em\_estoba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gp\_estl:\*  
11: gp\_estl2:\*  
12: gp\_hic:\*  
13: gp\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_lin:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69.4	56.4	530	10	AW761136 sl63f12.y
2	67.4	54.8	765	10	BE659470 GM700009B
3	66.2	53.8	764	10	AW439333 GM210007A
4	64.6	52.5	540	10	AW432561 sh75h06.y
5	62.2	50.6	497	11	BG551208 sad34e05.y
6	61.4	49.9	543	10	AW616232 EST307271
7	58.8	47.8	574	10	BE124753 EST393788
8	58.2	47.3	514	10	AW586761 EST18384
9	58.2	47.3	586	11	BE997991 EST429714
10	58.2	47.3	635	10	AW584723 N210838e
11	58.2	47.3	655	10	AW584724 N210839e
12	58.2	47.3	682	10	AW584948 N211110e

13	57.2	46.5	491	11	BG551127 sad33f01.
14	56.6	46.0	335	10	AI486129 EST244450
15	56.6	46.0	608	10	AW035325 EST280688
16	56.6	46.0	633	11	BI208519 EST526559
17	56.6	46.0	732	11	BG887479 EST513330
18	56.4	45.9	541	11	BG048538 OV1_14_H0
19	54.8	44.6	674	10	BE037301 MP19C11.M
20	54.2	44.1	959	10	BE039612 OC01G11.O
21	54	43.9	911	10	BE18763 SCL074.H0
22	53.4	43.4	511	10	AI987319 660003H05
23	53.4	43.4	675	11	BG605300 WHE2331.B
24	53.4	43.4	726	11	BG643885 EST512079
25	53	43.1	449	11	BG320550 ZMO3_10h1
26	52.8	42.9	672	10	AI731699 BNLG1105
27	51.8	42.1	194	10	BE123398 946001F06
28	51.8	42.1	588	10	AI759012 605085F11
29	51.8	42.1	672	10	AI726030 BNLG1139
30	50.8	41.3	612	10	BE639441 946033B11
31	50.4	41.0	862	11	BF275530 GA_Eb002
32	50.2	40.8	615	11	BF474495 WHE0844.G
33	50.2	40.8	699	10	AU101324 AU101324
34	50.2	40.8	861	11	BG300821 HYSMB001
35	49.6	40.3	420	10	AW459464 sh41e10.y
36	49.4	40.2	539	11	BF650924 NF098E01E
37	49.2	40.0	586	11	BG605406 WHE2329.E
38	49	39.8	418	10	AI443020 sa30h05.y
39	48.6	39.5	380	10	AI026312 L0-529T3
40	48.6	39.5	490	10	AU084649 AU084649
41	48.6	39.5	558	11	BG606610 WHE2958.F
42	48.6	39.5	658	10	AW448025 BRY_1261
43	48	39.0	594	10	BE336433 sp25e05.y
44	48	39.0	706	10	BE214475 HV_CEB000
45	47	38.2	444	10	BE519330 945023A05

#### ALIGNMENTS

RESULT 1  
AW761136  
LOCUS  
DEFINITION  
sl63f12.y1 Gm-cl027 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl027-6024 5' similar to TR:065462 065462 RECEPTOR LIKE PROTEIN ; mRNA sequence.  
ACCESSION  
AW761136  
VERSION  
AW761136.1 GI:7693038  
KEYWORDS  
SOURCE  
ORGANISM  
soybean.  
Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
REFERENCE  
1 (bases 1 to 530)  
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Rheising,B., Allen,M., Bowers,T., Person,B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schuck,R., Riller,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or



info@genomesystems.com web site: www.genomesystems.com  
 Insert Length: 1680 Std Error: 0.00  
 High quality sequence stop: 420.

## FEATURES

## SOURCE

1. 530  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl027-6024"  
 /clone\_lib="Gm-cl027"  
 /tissue\_type="cotyledons of 3- and 7-day-old Williams  
 seedlings"  
 /lab\_host="DH10B"

/note="Vector: pBluescript II SK+, Site\_1: EcoRI, Site\_2:  
 XhoI. This cDNA library was constructed from mRNA isolated  
 from cotyledons of 3- and 7-day-old Williams seedlings  
 which were propagated on paper towels with distilled  
 water. The cotyledons were flash-frozen in liquid  
 nitrogen, then lyophilized for 72 hours. Unequal amounts  
 of mRNA was used for cDNA synthesis. Stratagene's cDNA  
 Synthesis kit (catalog number 200401) was used to  
 synthesize the cDNA. First-stranded synthesis was  
 performed with 5-methyl dCTP, hence the ligated cDNA was  
 hemimethylated. A modification of Stratagene's  
 first-strand synthesis primer was used. An anchor  
 nucleotide (V-A, C, or G) was added to the 3' end of the  
 primer (GACAGACAGACAGACAGACTGCTCCAG(1)18) to anchor  
 the primer at the 5' end of the poly(A) tract. After  
 second-strand synthesis, the cDNA ends were filled in  
 with cloned Pfu DNA, ligated to EcoRI adapters and  
 subsequently phosphorylated. The XhoI site within the  
 first-strand synthesis primer was then restricted by  
 digestion with XhoI; all XhoI sites in the cDNA would be  
 protected by their hemimethylated status. The cDNA  
 constructs were size-fractionated with a 500 bp cutoff,  
 using GibcoBRL Life Technologies' cDNA Size Fractionation  
 column. The column eluent was then ligated into  
 Stratagene's pBluescript(II) SK+ predigested vector  
 (pBluescript II SK(+)) that has been digested with EcoRI  
 and XhoI, and phosphorylated by Stratagene). 97% of the  
 white and blue colonies appear to contain recombinant  
 plasmids with cDNA inserts, based on size (n=30). This  
 library was constructed by Dr. Paul Keim and Dr. Virginia  
 Corryell."

BASE COUNT 136 a 145 c 136 g 113 t  
 ORIGIN

Query Match 56.4%; Score 69.4; DB 10; Length 530;  
 Best Local Similarity 98.6%; Pred. No. 3.3e-12;  
 Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gttatgtaaatgtgagatgcttaacacctgagcctaagtgcgaatgcgaaggctg 63  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 407 GTATTGTAATGTGAGATGCTTACCAACCCGTATGACCTAATGCTCATTGCCAGGCGTG 466  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 64 cactgactgct 74  
 ||||||||||||  
 DB 467 CACTGACTGCT 477

RESULT 2  
 BE659470/c 765 bp mRNA EST 24-MAY-2001  
 LOCUS GM700009B10H8 Gm-r1070 Glycine max cDNA clone Gm-r1070-3447 3',  
 DEFINITION  
 mRNA sequence.

ACCESSION BE659470  
 VERSION BE659470.1 GI:9985362  
 KEYWORDS EST.

## ORGANISM

soybean.  
 Glycine max  
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;  
 Rosidae: eustoids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

REFERENCE 1 (bases 1 to 765)  
 Vodka, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V.,  
 Erpedding, J., Rapp, C., Shoop, E., Pardins, J., Liu, L. and Lewin, H.  
 A Functional Genomics Program for Soybean (NSF 9872565)  
 Unpublished (1999)  
 COMMENT

Other ESTs: AN432561 corresponding to Gm-cl015-5652 (5')  
 Contact: Vodka, L.O., PI, A Functional Genomics Program for  
 Soybean (NSF 9872565)  
 Lewin, H. A., Director, Keck Center for Comparative and Functional  
 Genomics  
 University of Illinois  
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
 Tel: (217) 244-6147  
 Fax: (217) 333-4582  
 Email: l-vodka@uiuc.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134. For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or info@genome  
 systems.com web site: www.genomesystems.com  
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

## FEATURES

## SOURCE

Location/Qualifiers

1. 765  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="Gm-r1070-3447"  
 /clone\_lib="Gm-r1070"

/note="The library Gm-r1070 is a sequence-driven, rerecked  
 set of 9,216 clones selected from cDNA libraries from  
 various tissues and stages of development of soybean that  
 represent 2,639 sequences from immature cotyledons, 1,770  
 from immature seed coats, 3,938 from flowers, and 869  
 from young pods. The 5' ESTs of the source clones from  
 the different libraries was used to select singletons, or  
 a representative of each contig, which were rerecked to  
 form library Gm-r1070. The cDNA clones of the rerecked  
 Gm-r1070 library were then sequenced at the 3' end. The  
 contig analysis to select unique genes was performed by  
 the laboratory of Ernest Retzel, Center for Computational  
 Genomics and Bioinformatics, University of Minnesota,  
 http://www.cbc.umn.edu/ResearchProjects/Soybean/Index.html  
 Rerecking was performed by Genome Systems, St. Louis,  
 http://www.genomesystems.com, and 3' sequencing by the  
 Keck Center for Comparative and Functional Genomics,  
 University of Illinois,  
 http://www.life.uiuc.edu/biotech/keck.html. Note: The  
 corresponding 5' EST from each clone in the Gm-r1070  
 library is listed in the 'OTHER EST' field. The detailed  
 information on the source library for each clone can also  
 be obtained by referring to the Genome Systems clone ID of  
 the original cDNA library that is also listed under  
 'OTHER EST'."

BASE COUNT 227 a 167 c 150 g 197 t 24 others  
 ORIGIN

Query Match 54.8%; Score 67.4; DB 10; Length 765;  
 Best Local Similarity 95.8%; Pred. No. 1.6e-11;  
 Matches 68; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gttatgtaaatgtgagatgcttaacacctgagcctaagtgcgaatgcgaaggctg 63  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 542 GTATTGTAATGTGAGATGCTTNNAAACCCGTATGACCTAATGCTGCAATCGAGGCGTG 483  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 64 cactgactgct 74  
 ||||||||||||  
 DB 482 CACTGACTGCT 472

RESULT 3  
 AM349333/c 764 bp mRNA EST 04-OCT-2000  
 LOCUS AM349333



DEFINITION GM210007A20A6 Gm-r1021 glycine max cDNA clone Gm-r1021-2316 3', mRNA sequence.

ACCESSION AM349333

VERSION AM349333.1 GI:6847043

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

REFERENCE 1 (bases 1 to 764)  
Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V., Erpelding, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L., and Lewin, H. A Functional Genomics Program for Soybean (NSF 9872565)

TITLE A Functional Genomics Program for Soybean (NSF 9872565)

JOURNAL Unpublished (1999)

COMMENT Other ESTs: A1661129  
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)  
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics  
University of Illinois  
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
Tel: (217) 244-6147  
Fax: (217) 333-4582  
Email: l-vodkin@uiuc.edu  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com  
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES  
source  
1..764  
/organism="Glycine max"  
/cultivar="Williams"  
/db\_xref="taxon:3847"  
/clone="Gm-r1021-2316"  
/clone\_lib="Gm-r1021"  
/tissue\_type="root"  
/lab\_host="X10-Gold"  
/note="Vector: pBluescript II XR, site\_1: EcoRI; site\_2: XhoI; library Gm-r1021 is a sequence-driven, rechecked set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stragene's cDNA synthesis kit (catalog #200401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.keim@naui.edu, virginia.coryell@naui.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota.  
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html  
Reracking was performed by Genome Systems, St. Louis.  
http://www.genomesystems.com and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois.  
http://www.life.uiuc.edu/biotech/keck.html."

BASE COUNT 239 a 161 c 151 g 193 t 20 others

ORIGIN

Query Match 53.8%; Score 66.2; DB 10; Length 764;  
Best local Similarity 95.8%; Pred. No. 3.8e-11;  
Matches 68; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gatttttaaatgagatgcttacaacctgctgacctaatgcaatgcaaggctg 63  
|||||  
566 gtaattgaattgacatgctttacacacctgattgacctaatgctgacgctg 507

QY 64 cactgactggt 74  
|||||  
Db 506 CACTGACTGCT 496

RESULT 4  
AM432561  
LOCUS 540 bp mRNA EST 18-JUL-2000

DEFINITION sh75n06.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl015-5652 5' similar to TR:065462 065462 RECEPTOR LIKE PROTEIN ; mRNA sequence.

ACCESSION AM432561

VERSION AM432561.1 GI:6963868

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

REFERENCE 1 (bases 1 to 540)  
Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna, A., Bolla, B., Matra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
Insert length: 1008 Std Error: 0.00  
Seq primer: -40RP from Gdbco  
High quality sequence stop: 380.

FEATURES  
source  
1..540  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-5652"  
/clone\_lib="Gm-cl015"  
/tissue\_type="Mature flowers, field grown plants"  
/lab\_host="X10-Gold"  
/note="Vector: pBluescript II XR, site\_1: EcoRI; site\_2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stragene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence followed by the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into X10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."

BASE COUNT 123 a 142 c 138 g 137 t

ORIGIN

Query Match 52.5%; Score 64.6; DB 10; Length 540;  
Best local Similarity 94.4%; Pred. No. 1.2e-10;  
Matches 67; Conservative 0; Mismatches 4; Indels 0; Gaps 0;



```

OY      4  gtattgtaattgagatgcttacacacctgatagccttaatgcgaatgcgaggctg 63
          |||||||
DB      468  GTATTGTAATGATGAGATGCTTACAAACCATGATGACTGATGCTGACGAGGCTG 527
OY      64  cactgactgct 74
          |||||||
DB      528  CACTGACTGCT 538

RESULT   5
BG551208
LOCUS    BG551208      497 bp      mRNA      EST      09-APR-2001
DEFINITION
sa034605.y1 Gm-cl074 Glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-cl074-2145 5' similar to TR:09SV14 09SV14 ES43 LIKE PROTEIN
; mRNA sequence.
ACCESSION BG551208
VERSION   BG551208.1 GI:13562988
KEYWORDS EST.
SOURCE   soybean.
ORGANISM Glycine max
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
           Glycine.
REFERENCE 1 (bases 1 to 497)
AUTHORS  Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
          ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,C., Beck,C.,
          Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
          ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
          ,R., Rutter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
          ,R., Waterston,R. and Wilson,R.
          Public Soybean EST Project
          Unpublished (1999)
          Contact: Shoemaker R/Public Soybean EST Project
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          This clone is available through: Genome Systems, Inc. 4633 World
          Parkway Circle St. Louis, Missouri 63134 for further information
          call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
          427-3324 or contact: clones@genomesystems.com or
          info@genomesystems.com web site: www.genomesystems.com
          High quality sequence stop: 380.
FEATURES
source   Location/Qualifiers
          1..497
          /organism="Glycine max"
          /db_xref="taxon:3847"
          /clone="GENOME SYSTEMS CLONE ID: Gm-cl074-2145"
          /clone_lib="Gm-cl074"
          /tissue_type="seedlings induced for HR (hypersensitive
          response)"
          /dev_stage="9-11 day old"
          /lab_host="DH10B"
          /note="Vector: pBluescript II SK(+). Site:1: EcoRI; Site:2:
          XhoI. The cDNA library was constructed from mRNA isolated
          from 9-11 day old seedlings that were induced for HR
          (hypersensitive response) by vacuum infiltrating plant
          tissue with Pseudomonas syringae pv. glycinea carrying the
          avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
          unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
          and 53 hrs after inoculation and their mRNA pooled equally
          for cDNA construction. The library was prepared using the
          Stratagene pBluescript II SK(+) library construction kit.
          Complementary DNA was synthesized from mRNA using a primer
          consisting of a poly(dT) sequence with an XhoI restriction
          site. EcoRI adaptors were ligated to the blunt-ended cDNA
          fragments followed by XhoI digestion. The cDNA insert is
          protected from XhoI digestion via methylation during first
          strand synthesis. The cDNA fragments were directionally

```

```

BASE COUNT  119 a      140 c      129 g      109 t
ORIGIN

Query Match      50.6%; Score 62.2; DB 11; Length 497;
Best Local Similarity 95.5%; Pred. No. 6.9e-10;
Matches 64; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      4  gtattgtaattgagatgcttacacacctgatagccttaatgcgaatgcgaggctg 63
          |||||||
DB      431  GTATTGTAATGATGAGATGCTTACAAACCATGATGACTGATGCTGACGAGGCTG 490
OY      64  cactgactg 70
          |||||||
DB      491  CACTGACT 497

RESULT   6
AM616232
LOCUS    AM616232      543 bp      mRNA      EST      18-MAY-2001
DEFINITION
EST307271 L. hirsutum trichome, Cornell University Lycopersicon
hirsutum cDNA clone CLHT1123 5', mRNA sequence.
ACCESSION AM616232
VERSION   AM616232.1 GI:7322266
KEYWORDS EST.
SOURCE   Lycopersicon hirsutum.
ORGANISM Lycopersicon hirsutum.
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
           Lycopersicon.
REFERENCE 1 (bases 1 to 543)
AUTHORS  van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
          ,F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W.,
          Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
          Generation of ESTs from wild tomato (Lycopersicon hirsutum)
          trichomes
          Unpublished (2000)
          Contact: CUGI
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall, Clemson, SC 29634, USA
          Email: http://www.genome.clemson.edu/orders/index.html
          5 prime sequence.
FEATURES
source   Location/Qualifiers
          1..543
          /organism="Lycopersicon hirsutum"
          /db_xref="taxon:62890"
          /clone="CLHT1123"
          /clone_lib="L. hirsutum trichome, Cornell University"
          /tissue_type="trichome"
          /dev_stage="mixed stages"
          /note="Leaves of various stages were shaken in liquid
          nitrogen, shearing off trichomes. This procedure yielded a
          mixture of cells which is highly enriched for trichome
          likely with minor contaminations of other types of leaf
          cells"
BASE COUNT  139 a      134 c      133 g      137 t
ORIGIN

Query Match      49.9%; Score 61.4; DB 10; Length 543;
Best Local Similarity 86.1%; Pred. No. 1.3e-09;
Matches 68; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY      4  gtattgtaattgagatgcttacacacctgatagccttaatgcgaatgcgaggctg 63
          |||||||

```



Db	456	GTAACGTAAATGTAAGTCCATACAAATCCTATGACCACTCATGGTTTCATGTGAAGGCCTG	515
Qy	64	cactgactggtgaattcca	82
Db	516	CAGTGACTGTGTAAGTCTGA	534
RESULT	7		
LOCUS	BE124753	574 bp	mRNA EST 07-SEP-2000
DEFINITION	EST393188 GVN Medicago truncatula cDNA clone pGVN-67F7, mRNA sequence.		
ACCESSION	BE124753		
VERSION	BE124753.1	GI:8529310	
KEYWORDS	EST.		
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;		
	Medicago.		
REFERENCE	1 (bases 1 to 574)		
AUTHORS	Fedorova,M., Plesnon,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.		
TITLE	ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Carroll P. Vance Department of Agronomy and Plant Genetics University of Minnesota 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA Tel.: 612 625 5715 Fax: 651 649 5058 Email: vance004@maroon.tc.umn.edu University of Minnesota name:M261621e TIGR sequence name:MTCCC28TK More information is available at: <a href="http://chrysie.camu.edu/medicago">http://chrysie.camu.edu/medicago</a> Seq primer: SKmod (CTA GAA CTA gtc gat CC).		
FEATURES	Location/Qualifiers		
source	1..574		
	/organism="Medicago truncatula"		
	/cultivar="genotype A17"		
	/db_xref="taxon:3880"		
	/clone="pgVN-67F7"		
	/clone_lib="GVN"		
	/tissue_type="N2-fixing root nodules"		
	/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"		
	/lab_host="E. coli strain XLOLR"		
	/note="Vector: pluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the uni-zap XR vector from stragene and packaged using gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-Assist helper phage and propagated in XLOLR cells."		
BASE COUNT	155 a 114 c 137 g 167 t	1 others	
ORIGIN			
Query Match	47.8%	Score 58.8:	DB 10; Length 574;
Best Local Similarity	88.7%;	Pred. No. 8.9e-09;	
Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;			
Qy	4	gtatgtataatgtgagaaccttaaacccgatctaactaatggcgcaatgcagagctg	63
Db	439	GTAATGTAATAGTAGAGTCCCTACAAACCCTGATGATCTGATGATTTCANTGTGAGGCGCTG	498
Qy	64	cactgactggtgt	74

DB	499	CAGTCACTGCT	509	
RESULT	8			
LOCUS	AM586761			
DEFINITION	ESJ318384 MHAM Medicago truncatula/Glomus versiforme mixed EST library cDNA clone pMHAM-55G12, mRNA sequence.			
ACCESSION	AM586761			
VERSION	AM586761.1			
KEYWORDS	GI:7266275			
SOURCE	EST.			
ORGANISM	Medicago truncatula/Glomus versiforme mixed EST library. Medicago truncatula/Glomus versiforme mixed EST library. Eukaryota; mixed EST libraries.			
REFERENCE	1 (bases 1 to 514)			
AUTHORS	Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.			
TITLE	ESTs from roots of Medicago truncatula after colonization with Glomus versiforme			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Maria J. Harrison Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73401, USA Tel: 580-223-5810 Fax: 580-223-7380 Email: mjharrison@noble.org Noble EST name:N25126e TIGR sequence name:MTDAV429K More information is available at. 'http://chrstyle.tamu.edu/medicago Seq primer: SKmod (CTA GAA CTA gtc gat CC). Location/Qualifiers 1..514 /organism="Medicago truncatula/Glomus versiforme mixed EST library" /cultivar="Medicago truncatula genotype A17" /db_xref="taxon:119092" /clone="pMHAM-55G12" /clone_id="MHAM" /tissue.type="roots colonized with Glomus versiforme" /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages." /lab_host="E. coli strain XL0LR" /note="Vector: pBluescript SK-. Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the unzip XR vector from stratagene and packaged using gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in XL0LR cells."			
BASE COUNT	135 a			
ORIGIN	101 c			
	127 g			
	151 t			
Query Match	47.3%			
Best Local Similarity	88.7%			
Matches	63; Conservative			
	0; Mismatches			
	8; Indels			
	0; Gaps			
	0;			
OY	4	gtatgttaagtgaagccttaaacccgtagtactaataagcgcaatgagagctg	63	
DB	415	gtattgttaaatgttgagatgacctacacccctgattgattctgacgtgagggctg	474	
OY	64	cactgactgctg	74	
DB	475	cagtgactgctg	485	



	BE997991	586 bp	mRNA	EST	06-OCT-2000
LOCUS	BE997991				
DEFINITION	EST#29714 GVSN Medicago truncatula cDNA clone pgVSN-BF9, mRNA sequence.				
ACCESSION	BE997991				
VERSION	BE997991.1	GI:10698267			
KEYWORDS	EST.				
SOURCE	barrrel medic.				
ORGANISM	Medicago truncatula				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.				
REFERENCE	1 (bases 1 to 586)				
AUTHORS	Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Townm,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.				
TITLE	ESTs from senescent nodules of Medicago truncatula				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Carroll P. Vance Department of Agronomy and Plant Genetics University of Minnesota 411 Borlang Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA Tel: 612 625 5715 Fax: 651-649-5058 Email: vance004@maroon.tc.umn.edu University of Minnesota name: M271969e TIGR sequence name: MRAK29TK More Information is available at: <a href="http://chrysie.tamu.edu/medicago">http://chrysie.tamu.edu/medicago</a> Seq primer: SKmod CTA gaa cta ggc gat cc). Location/Qualifiers 1..586 /organism="Medicago truncatula" /cultivar="genotype A17" /db_xref="taxon:3880" /clone="pgVSN-BF9" /clone_id="GVSN" /tissue.type="senescent root nodules" /der_stage="mixture of effective nodules from 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage" /lab_host="E. coli strain SOLR" /note="Vector: plunscript SK +/-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of effective nodules of 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA insets were excised from the recombinant lambda-ZAP phage using Ex-assist helper phage and propagated in SOLR cells."				
BASE COUNT	158 a 114 c 143 g 171 t				
ORIGIN					
	Query Match 47.3%; Score 58.2; DB 11; Length 586; Best Local Similarity 88.7%; Pred. No.1.4e-08; Matches 63; Conservative 0; Matches 8; Indels 0; Gaps 0;				
OY	4 gtattgaagtgtgagcctcaaccggatgactaataggcgcatcgagagctg 63 				
Dd	432 GTATTGTAAAGTGTGAGATGCCCTCAACACCCTCATGATCTGAGTGTCAGTGAGGGCTG 491 				
OY	64 cactgactgctt 74 				
Dd	492 CAGTGACTGCTG 502				
RESULT 10					
LOCUS	AM584723/C	635 bp	mRNA	EST	07-SEP-2000
DEFINITION	N210838e MHAM Medicago truncatula/Glommus versiforme mixed EST library cDNA clone MHAM-K19, mRNA sequence.				

ACCESSION	AM584723
VERSION	AM584723.1 GI:7261777
KEYWORDS	EST.
SOURCE	Medicago truncatula/Glomsus versiforme mixed EST library.
ORGANISM	Eukaryota; mixed EST libraries.
REFERENCE	Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M. ESTs from roots of Medicago truncatula after colonization with Glomsus versiforme
AUTHORS	Unpublished (2000)
TITLE	Contact: Harrison M.J.
JOURNAL	Plant Biology Division
COMMENT	The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73401 Tel: 580-223-5810 Fax: 580-221-7380 Email: mjharrison@noble.org Other name: MHAM-7A-F10; Date: 3/14/00; Updated to the database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at ' <a href="http://chrysis.tamu.edu/medicago/">http://chrysis.tamu.edu/medicago/</a> '. Seq primer: '93.  Location/Qualifiers 1..635 /organism="Medicago truncatula/Glomsus versiforme mixed EST library" /cultivar="Medicago truncatula genotype A17" /db_xref="taxon:119092" /clone="MHAM-7K19" /clone_lib="MHAM" /tissue_type="Roots colonized with Glomsus versiforme" /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomsus versiforme. The library was made from a mixture of RNA from each of these stages." /lab_host="E. coli strain XLOLR" /note="Vector: pBluescript SK-, Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomsus versiforme. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."
BASE COUNT	202 a 159 c 103 g 171 t
ORIGIN	
Query Match	47.3%; Score 58.2; DB 10; Length 635;
Best local Similarity	88.7%; Pred. No. 1.4e-08;
Matches	63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY	4 gttattgaatgtgagatgccttaaccctgatgacctaatggcgaatggaaggcgctg 63       Db 468 GTATTGTAATGTGAGATGCCCTTACMACCCGTGATGATCTGATGGTTTCAGTGAGGCGTG 409       OY 64 cactgactgagt 74          Db 408 CACTGACTGCT 398
RESULT 11	
LOCUS	AM584724/c
DEFINITION	N210839e MHAM Medicago truncatula/Glomsus versiforme mixed EST
ACCESSION	AM584724
VERSION	AM584724.1 GI:7261778
KEYWORDS	EST.
SOURCE	Medicago truncatula/Glomsus versiforme mixed EST library.
ORGANISM	Medicago truncatula/Glomsus versiforme mixed EST library. Eukaryota; mixed EST libraries.



REFERENCE	1 (bases 1 to 655)	
AUTHORS	Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.	
TITLE	ESTs from roots of Medicago truncatula after colonization with Glomus versiforme	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Harrison M.J. Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73401 Tel: 580-223-5810 Fax: 580-221-7380 Email: mjharrison@noble.org Other name: MHAM-7a-F11; Date: 3/14/00; updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at 'http://chrystie.tamu.edu/medicago'. Seq primer: T3.	
FEATURES	Location/Qualifiers	
SOURCE	1..655	
	/organism="Medicago truncatula/Glomus versiforme mixed EST library"	
	/cultivar="Medicago truncatula genotype A17"	
	/db_xref="taxon:119092"	
	/clone="MHAM-7K21"	
	/clone_lib="MHAM"	
	/tissue_type="roots colonized with Glomus versiforme"	
	/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."	
	/lab_host="E. coli strain XL0LR"	
	/note="Vector: pluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the UniZap XR vector from Stratagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper-phage and propagated in XL0LR cells."	
BASE COUNT	207 a 163 c 108 g 177 t	
ORIGIN		
Query Match	47.3%; Score 58.2; DB 10; Length 655;	
Best Local Similarity	88.7%; Pred. No. 1.4e-08;	
Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;		
OY	4 gttatgtcaatgtagagatgccttaacaccctgataactaagcgcaatgcgagagctg 63	
Db	468 gTATTGTAAAGTGTGAGATGCCCTACACCCGTGATGATCTGATGATGTGAGAGGCGTG 409	
OY	64 cactgactgct 74	
Db	408 CAGTGACTGCT 398	
RESULT 12		
AW584948/c		
LOCUS	AW584948 682 bp mRNA EST 07-SEP-2000	
DEFINITION	N211110e MHAM Medicago truncatula/Glomus versiforme mixed EST library cDNA clone MHAM-7J12, mRNA sequence.	
ACCESSION	AW584948	
VERSION	AW584948.1 GI:7262002	
KEYWORDS	EST.	
SOURCE	Medicago truncatula/Glomus versiforme mixed EST library.	
ORGANISM	Medicago truncatula/Glomus versiforme mixed EST library	
REFERENCE	1 (bases 1 to 682)	
AUTHORS	Eukaryota, mixed EST libraries.	
TITLE	Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.	
JOURNAL	ESTs from roots of Medicago truncatula after colonization with Glomus versiforme	
COMMENT	Unpublished (2000)	

COMMENT	
Contact: Harrison M.J. Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73401 Tel: 580-223-5810 Fax: 580-221-7380 Email: mjharrison@noble.org Other name: MHAM-76-E06; Date: 3/14/00; Updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at ' <a href="http://chrystle.tamv.edu/medicago/">http://chrystle.tamv.edu/medicago/</a> '. Seq primer: T3.	
FEATURES SOURCE	Location/Qualifiers
	1..682
	/organism="Medicago truncatula/Glommus versiforme mixed EST library"
	/cultivar="Medicago truncatula genotype A17"
	/db_xref="taxon:119092"
	/clone="MHAM-7J12"
	/clone_lib="MHAM"
	/tissue_type="roots colonized with Glommus versiforme"
	/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glommus versiforme. The library was made from a mixture of RNA from each of these stages."
	/lab_host="E. coli strain XLOLR"
	/note="Vector: plusescript SK-. Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glommus versiforme. The cDNA was directionally ligated into the Unizap XR vector from Stragene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."
BASE COUNT ORIGIN	212 a 172 c 115 g 183 t
Query Match	47.3%; Score 58.2; DB 10; Length 682;
Best Local Similarity	88.7%; Pred. No. 1.4e-08;
Matches	63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY 4 gtattgtaaatgttagatgccttaacaccgcgatgactaaatggcgaatcgaggctg 63       	
Dh 468 GTATTGTAAGTGTGAGATGCCCTACMACCGTAGTGATCTGATGGTTGAGAGGCTG 409       	
OY 64 cactgacctgct 74 	
Dh 408 CAGTGACTGCT 398	
RESULT 13	
BGS51127	
LOCUS DEFINITION	BGS51127 491 bp mRNA EST 09-APR-2001
ACCESION VERSION KEYWORDS	sad33f01.y1 Gm-cl074 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl074-2162 5' similar to TR:O9SV14 Q9SV14 ES43 LIKE PROTEIN /, mRNA sequence.
ORGANISM SOURCE	BGS51127 BGS51127.1 GI:13562907 EST.
	soybean.
	Glycine max
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosoids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE AUTHORS	1 (bases 1 to 491) Shoemaker,R., Keim,P., Vodkin,L., Erpellding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepien,M., Thelning,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R. Public Soybean EST Project



JOURNAL Unpublished (1999)  
COMMENT Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
High quality sequence stop: 421.

## FEATURES

source  
1. 491  
Location/Qualifiers  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1074-2162"  
/clone\_lib="Gm-c1074"  
/tissue\_type="seedlings induced for HR (hypersensitive response)"  
/dev\_stage="9-11 day old"  
/lab\_host="DH10B"  
/note="Vector: pBluescript II SK+, Site.1: EcoRI; Site.2: XhoI; The cDNA library was constructed from mRNA isolated from 9-11 day old seedlings that were induced for HR (hypersensitive response) by vacuum infiltrating plant tissue with *Pseudomonas syringae* pv. *glycinea* carrying the avr gene (Genetics 141:1597-1604). Plant tissue (expanded unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36, and 53 hrs after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene pBluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into *E. coli* Electromax DH10B host cells. Plant care, inoculations, and library construction were performed by Steve Clough (Lila Vodkin lab, University of Illinois)."

BASE COUNT 117 a 138 c 129 g 107 t  
ORIGIN

Query Match 46.5%; Score 57.2; DB 11; Length 491;  
Best Local Similarity 95.2%; Pred. No. 2.8e-08;  
Matches 59; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ggtatgtgaatgtgagatgcttaaacctgtagcttaatggcgcaatgcgaggtc 62  
|||||  
Db 430 ggtattgtaaatgtgagatgcttaaacctgtagcttaatggcgcaatgcgaggtc 489  
|||||

QY 63 gc 64  
||  
Db 490 gc 491

RESULT 14  
AI486129  
LOCUS AI486129 335 bp mRNA EST 18-MAY-2001  
DEFINITION EST244450 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
AI486129  
ACCESSION AI486129  
VERSION AI486129.1 GI:4381500  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum

## REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanales; Solanales; Solanum; Lycopersicon.

## AUTHORS

1 (bases 1 to 335)  
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.

## JOURNAL

## COMMENT

Unpublished (1999)  
Generation of ESTs from tomato carpel tissue  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.

## FEATURES

source  
1. 335  
Location/Qualifiers  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLEDSK16"  
/clone\_lib="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="XL1-Blue MRF"  
/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2: XhoI; cLED - Tomato Carpel EST Library. Oligo(dT)-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 99 a 76 c 72 g 88 t  
ORIGIN

Query Match 46.0%; Score 56.6; DB 10; Length 335;  
Best Local Similarity 87.3%; Pred. No. 4.1e-08;  
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 gttatgtaattgagatgcttaaacctgtagcttaatggcgcaatgcgaggtc 63  
|||||  
Db 168 gttacttaattgtaaatgtgagatgcttaaacctgtagcttaatggcgcaatgcgaggtc 227  
|||||

QY 64 cactgactgt 74  
|||||  
Db 228 CAGTGACTGCT 238

## RESULT 15

AW035325  
LOCUS AW035325 608 bp mRNA EST 18-MAY-2001  
DEFINITION EST280688 tomato callus, TAMU Lycopersicon esculentum cDNA clone  
CLEC40312 similar to A. thaliana ESA3-like protein, mRNA sequence.

ACCESSION AW035325  
VERSION AW035325.1 GI:5894081  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanales; Solanales; Solanum; Lycopersicon.

## REFERENCE

1 (bases 1 to 608)  
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.

## JOURNAL

## COMMENT

Unpublished (1999)  
Generation of ESTs from tomato callus tissue  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>



5 prime sequence.  
FEATURES  
Location/Qualifiers  
source  
1..608  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone\_lib="CLEC40J12"  
/clone\_lib="tomato callus, TAMU"  
/tissue\_type="callus"  
/dev\_stage="25-40 days old"  
/lab\_host="XLI-Blue MRP"  
/note="Vector: pBluescript SK(-); site\_1: EcoRI; site\_2: XhoI; supplier: Giovannoni laboratory; CLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato Callus EST Library"  
BASE COUNT  
158 a 149 c 147 g 154 t  
ORIGIN

Query Match 46.0%; Score 56.6; DB 10; Length 608;  
Best Local Similarity 87.3%; Pred. No. 4.6e-08;  
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 gttattgaatgtgagatgccttaaacacctgtagcttaattgagcgaatgcgagagctg 63  
|||||  
Db 468 GTACTGTAAATGTGAATGCCATACATCCTGATGACCTCATGTTCAATGTGAGGCGT 527

QY 64 cactgactggt 74  
|||  
Db 528 CAGTGAAGTGT 538

RESULT 16  
BI208519 633 bp mRNA EST 11-JUL-2001  
LOCUS  
DEFINITION  
EST526559 CTO5 Lycopersicon esculentum cDNA clone CTO51768 5' end,  
mRNA sequence.  
ACCESSION  
BI208519 GI:14686243  
VERSION  
BI208519.1  
KEYWORDS  
EST.  
SOURCE  
tomato.  
ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 633)  
van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,  
Romning,C. and Tanksley,S.  
Generation of ESTs from Tomato Suspension Cultures  
Unpublished (2001)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.  
Location/Qualifiers  
1..633  
/organism="Lycopersicon esculentum"  
/cultivar="TA496, E6203"  
/db\_xref="taxon:4081"  
/clone\_lib="CTOS1768"  
/clone\_lib="CTOS"  
/tissue\_type="suspension cultures"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; Suspension cultures of L. esculentum E6203 were grown in Murashige and Skoog based medium, supplemented with 15% coconut milk (filter sterilized and added after autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8). Fresh medium was added every 7 days, and cultures were

grown at 25 C, with 12hrs of light and continuous shaking."  
BASE COUNT  
178 a 134 c 149 g 172 t  
ORIGIN

Query Match 46.0%; Score 56.6; DB 11; Length 633;  
Best Local Similarity 87.3%; Pred. No. 4.7e-08;  
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 gttattgaatgtgagatgccttaaacacctgtagcttaattgagcgaatgcgagagctg 63  
|||||  
Db 226 GTACTGTAAATGTGAATGCCATACATCCTGATGACCTCATGTTCAATGTGAGGCGT 285

QY 64 cactgactggt 74  
|||  
Db 286 CAGTGAAGTGT 296

RESULT 17  
BG887479 732 bp mRNA EST 30-MAY-2001  
LOCUS  
DEFINITION  
EST513330 CSTD Solanum tuberosum cDNA clone CSTD5N3 5' sequence,  
mRNA sequence.  
ACCESSION  
BG887479 GI:14264565  
VERSION  
BG887479.1  
KEYWORDS  
EST.  
SOURCE  
potato.  
ORGANISM  
Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 732)  
van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemingo,A.,  
Bouqri,O., Buell,C.R., Romning,C., Tanksley,S. and Baker,B.  
Generations of ESTs from dormant potato tubers  
Unpublished (2001)  
Contact: Cathy Romning  
The Institute for Genomic Research  
For clone info: please contact Research Genetics, Libraries  
division tel 1-800-711-6195, email [cdna@resgen.com](mailto:cdna@resgen.com)  
Seq primer: M13F-R.  
Location/Qualifiers  
1..732  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone\_lib="CSTD5N3"  
/clone\_lib="CSTD"  
/tissue\_type="dormant tuber"  
/dev\_stage="one month post-harvest"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); site\_1: EcoRI; site\_2: XhoI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tuber, avoiding the buds and epidermis. Tubers were stored for one month post-harvest at 40C. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as p4 in Tanksley lab notebooks."  
BASE COUNT  
191 a 181 c 169 g 191 t  
ORIGIN

Query Match 46.0%; Score 56.6; DB 11; Length 732;  
Best Local Similarity 87.3%; Pred. No. 4.8e-08;  
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 gttattgaatgtgagatgccttaaacacctgtagcttaattgagcgaatgcgagagctg 63  
|||||  
Db 511 GTACTGTAAATGTGAATGCCATACATCCTGATGACCTCATGTTCAATGTGAGGCGT 570

QY 64 cactgactggt 74



```

Db      571 CAGTCACTGTGT 581
|||||
RESULT 18
LOCUS   BG048538
DEFINITION OVI_14-H02.g2_A002 Ovary 1 (OVI) Sorghum bicolor cDNA, mRNA
ACCESSION BG048538
VERSION   BG048538.1
KEYWORDS  EST.
SOURCE    sorghum.
ORGANISM  Sorghum bicolor
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
           clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 541)
AUTHORS  Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
           ,L.H.
TITLE     An EST database from Sorghum: ovaries of varying immature stages
JOURNAL   Unpublished (2000)
COMMENT   Contact: Cordonnier-Pratt MM
           Department of Botany
           The University of Georgia
           Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
           Tel: 706 542 1860
           Fax: 706 542 1805
           Email: mmp@prattuga.edu
           Sequences have been trimmed to exclude polyA, vector and regions
           below Phred quality 16. The threshold for highest quality sequence
           is 20.
           Seq primer: PolyTMix
           High quality sequence start: 30
           High quality sequence stop: 541
           POLYA-No.

FEATURES
SOURCE    Location/Qualifiers
           1..541
           /organism="Sorghum bicolor"
           /db_xref="taxon:4558"
           /clone_lib="Ovary 1 (OVI)"
           /note="Organ: Mix of ovaries of varying immature stages
           from 8-week-old plants; Vector: pBluescript II from Lambda
           Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
           from poly-A RNA in the cloning vector Lambda Zap II.
           Clones to be sequenced were prepared by mass excision."
BASE COUNT 139 a 111 c 141 g 150 t
ORIGIN
Query Match 45.9%; Score 56.4; DB 11; Length 541;
Best Local Similarity 76.7%; Pred. No. 5,3e-08;
Matches 69; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Oy      4 gatttgtaagtgtgagatgccttaacacctgataacctaatggcgcaatggcgagcg 63
|||||
Db      52 GTATTGCAAAAGTGTGAGATGCCATACACCCCTGATGACCTCATGTCGACGCTTG 111
|||||
Oy      64 cactgactgctgaagtcttcacatgtaaggt 93
|||||
Db      112 CAAGGACTGTTCATCCATCTTCATGTG 141
|||||

RESULT 19
LOCUS   BE037301
DEFINITION MP19C11 MP Mesembryanthemum crystallinum cDNA 5' similar to es43
           like protein, mRNA sequence.
ACCESSION BE037301
VERSION   BE037301.1
KEYWORDS  EST.
SOURCE    common ice plant.
ORGANISM  Mesembryanthemum crystallinum

```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Aizaceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 674)
AUTHORS  Bohnerl,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
           ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
           Scara,G., Wheeler,M. and Zepeda,G.R.
           Functional Genomics of Plant Stress Tolerance
TITLE     Unpublished (2000)
JOURNAL   Unpublished (2000)
COMMENT   Contact: Michalowski,C.B.
           University of Arizona
           Bio Sciences West room 513, Tucson, AZ 85721, USA
           Tel: 520-621-7982
           Fax: 520-621-1697
           Email: cbm@u.arizona.edu

FEATURES
SOURCE    Location/Qualifiers
           1..674
           /organism="Mesembryanthemum crystallinum"
           /db_xref="taxon:3544"
           /clone_lib="MP"
           /tissue_type="apical meristem and leaf primordia"
           /dev_stage="6 weeks"
           /note="3 d 500mM NaCl"
BASE COUNT 171 a 184 c 164 g 155 t
ORIGIN
Query Match 44.6%; Score 54.8; DB 10; Length 674;
Best Local Similarity 83.8%; Pred. No. 1.8e-07;
Matches 62; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy      1 caggtatgttaagtgtgagatgccttaacacctgataacctaatggcgcaatggcgag 60
|||||
Db      545 CGGTACTGTAAATGTGAGATGCCCTACATCCTGATGACTTGATGAGTTCAATGTAGGCG 604
|||||
Oy      61 ctgactgactgtgt 74
|||||
Db      605 ATGCAATGATGTGT 618
|||||

RESULT 20
LOCUS   BE039612
DEFINITION OC01G11 OC Oryza sativa cDNA 5' similar to receptor like protein,
           mRNA sequence.
ACCESSION BE039612
VERSION   BE039612.1
KEYWORDS  EST.
SOURCE    Oryza sativa.
ORGANISM  Oryza sativa
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
           Echaroidae; Oryzae; Oryza.
REFERENCE 1 (bases 1 to 959)
AUTHORS  Bohnerl,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
           ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
           Scara,G., Wheeler,M. and Zepeda,G.R.
           Functional Genomics of Plant Stress Tolerance
TITLE     Unpublished (2000)
JOURNAL   Unpublished (2000)
COMMENT   Contact: Michalowski,C.B.
           University of Arizona
           Bio Sciences West room 513, Tucson, AZ 85721, USA
           Tel: 520-621-7982
           Fax: 520-621-1697
           Email: cbm@u.arizona.edu
           Insert Length: 1 Std Error: 0.00.

FEATURES
SOURCE    Location/Qualifiers
           1..959
           /organism="Oryza sativa"
           /strain="pokkali"
           /db_xref="taxon:4530"
           /clone_lib="OC"
           /tissue_type="roots"

```



BASE COUNT 279 a 225 c 271 g 180 t 4 others  
 ORIGIN

Query Match 44.1%; Score 54.2; DB 10; Length 959;  
 Best Local Similarity 73.9%; Pred. No. 3e-07;  
 Matches 68; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 4 gatttgtaagtgtgagatgcttaaaccttgatgacctgaatgagcgcaatgagagctg 63

DB 79 GTACTGCAAGTGGAGATGCCGTATTAACCGGATGATCTAATAGTGCACAGTGCAGGAGG 138

OY 64 cactgactgtgaattcacatgtaagtcgat 95

DB 139 CAAGGACTGTGTCCTCCCTTGATGGAAT 170

RESULT 21  
 BE418763 911 bp mRNA EST 24-JUL-2000

LOCUS SCL074.H01R90724 ITFC SCL Wheat Leaf Library Triticum aestivum  
 DEFINITION CDNA clone SCL074.H01, mRNA sequence.

ACCESSION BE418763

VERSION BE418763.1 GI:9416609

KEYWORDS

SOURCE

ORGANISM

bread wheat.  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
 ; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 911)  
 Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier,  
 S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,  
 Hermann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,  
 Langridge, P., Lazo, G.R., Lin, J., McGuire, P., Ogihara, Y.,  
 Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,  
 Sorrells, M., Warburton, M., and Wenzel, G.

International Triticale EST Cooperative (ITFC): Production of  
 Expressed Sequence Tags for Species of the Triticeae  
 Unpublished (2000)

JOURNAL Contact: Cloutier S  
 Cereal Research Centre, Agriculture & Agri-Foods Canada  
 Winnipeg, MB, Canada  
 Tel: 204 983 2340  
 Fax: 204 983 4604

Email: scloutier@em.agr.ca  
 International Triticale EST Cooperative (ITFC)  
 http://wheat.pw.usda.gov/genome.

FEATURES  
 source location/Qualifiers

1..911  
 /organism="Triticum aestivum"  
 /cultivar="Thatcher Lr1"  
 /db\_xref="taxon:4565"  
 /clone="SCL074.H01"  
 /clone\_lib="ITFC SCL Wheat Leaf Library"  
 /tissue\_type="young leaf"  
 /dev\_stage="after 24 hour challenge with LR-AVR1"  
 /note="Vector: Lambda Zap; 1.0 kbp average insert size."  
 BASE COUNT 215 a 231 c 248 g 175 t 42 others  
 ORIGIN

Query Match 43.9%; Score 54; DB 10; Length 911;  
 Best Local Similarity 79.7%; Pred. No. 3.5e-07;  
 Matches 63; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 4 gatttgtaagtgtgagatgcttaaaccttgatgacctgaatgagcgcaatgagagctg 63

DB 481 GTACTGCAAGTGGAGATGCCGTATTAACCGGATGATCTAATAGTGCACAGTGCAGGAGG 540

OY 64 cactgactgtgaattcacatgtaagtcgat 82

DB 541 CAAGGACTGTTCATCA 559

RESULT 22

LOCUS A1987319/c 511 bp mRNA EST 01-SEP-1999

DEFINITION 66003H05.x1 660 - Mixed stages of anther and pollen Zea mays cDNA,  
 mRNA sequence.

ACCESSION A1987319

VERSION A1987319.1 GI:5816403

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 511)  
 Maize ESTs from various cDNA libraries sequenced at Stanford  
 University

AUTHORS Unpublished (1999)

JOURNAL

Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221

Email: walbot@stanford.edu  
 Plate: 660003 row: H column: 05.

FEATURES  
 source location/Qualifiers

1..511  
 /organism="Zea mays"  
 /cultivar="Ohio43"  
 /db\_xref="taxon:4577"  
 /clone\_lib="660 - Mixed stages of anther and pollen"  
 /tissue\_type="whole premeiotic anthers to pollen shed"  
 /dev\_stage="premeiotic anthers to pollen shed"  
 /lab\_host="X10LR"  
 /note="Organ: anthers; Vector: Lambda Zap; Site: 1; Ecore: 1;  
 Site: 2; XhoI; Anther and pollen cDNA library.  
 Directionally sequenced with 5' end at the EcoRI site.  
 Created by Amie Franklin."

Query Match 43.4%; Score 53.4; DB 10; Length 511;  
 Best Local Similarity 79.7%; Pred. No. 4.9e-07;  
 Matches 63; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 4 gatttgtaagtgtgagatgcttaaaccttgatgacctgaatgagcgcaatgagagctg 63

DB 417 GTATTGCAAGTGGAGATGCCATCAACCGGATGATCTAATAGTGCACAGTGCAGGAGG 358

OY 64 cactgactgtgaattcacatgtaagtcgat 82

DB 357 CAAGGACTGTTCATCA 339

RESULT 23  
 LOCUS BG605300 675 bp mRNA EST 16-APR-2001

DEFINITION WHE2331.B03.D05ZS Wheat pre-anthesis spike cDNA library Triticum  
 aestivum CDNA clone WHE2331.B03.D05, mRNA sequence.

ACCESSION BG605300

VERSION BG605300.1 GI:13635303

KEYWORDS

SOURCE

ORGANISM

bread wheat.  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
 ; Triticeae; Triticum.







VERSION	A1759012.1	GI:5152847
KEYWORDS	EST.	
SOURCE	Zea mays.	
ORGANISM	Zea mays	

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta, Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.



REFERENCE	1 (bases 1 to 588)
AUTHORS	Walbot, V
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL	University
COMMENT	Unpublished (1999)
	Contact: Walbot V

FEATURES	Location/Qualifiers
source	1. .588

BASE COUNT	161 a	163 c	122 g	142 t
ORIGIN				

BASE COUNT	ORIGIN
174 a	162 c 145 g 190 t 1 others

BASE COUNT	134 a	176 c	186 g	116 t
ORIGIN				



















```

BASE COUNT      102 a      69 c      96 g      113 t
ORIGIN
      , 0 hours NaCl treatment"
      /tissue_type="leaf"
      /dev_stage="Six week old"
      /note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site_1:
      EcoRI; Site_2: XhoI"

```

Query Match	39.5%;	Score 48.6;	DB 10;	Length 380;
Best Local Similarity	80.3%;	Pred. No. 1.6e-05;		
Matches 57; Conservative	0;	Mismatches 14;	Indels 0;	Gaps 0;

```

Oy      4  gtaattgaatgtgagatgccttacaacccgtatgacccaatgycgcacatcgcagagctg 63
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      141  gtaattgcaagtgtgaatgccatataaacccggatgatctttatggtcacaattgtagaggttg 200

```

Qy	64	cactgactgt	74
Db	201	TAAGGACTGGT	211

RESULT	40
AU084649	
LOCUS	490 bp mRNA
DEFINITION	EST
	01-SEP-2000
	AU084649 Cryptomeria japonica inner bark Cryptomeria japonica cDNA
	clone CC1649 5', mRNA sequence.

VERSION	AU084649.1	GI:9962632
KEYWORDS	EST.	
SOURCE	Japanese cedar.	

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Spermatophyta; Coniferopsida; Coniferales; Cupressaceae;  
Cryptomeria.

AUTHORS	TITLE
Ujino-Ihara, T., Yoshimura, K., Ugawa, Y., Yoshimaru, H., Nagasaka, K. and Tsumura, Y.	Expression analysis of ESTs derived from the inner bark of

JOURNAL Plant Mol. Biol. 43, 451-457 (2000)  
COMMENT Contact: Tokuko Ujino-Ihara

Forestry and Forest Products Research Institute  
Matsumoto 1, Kuzizaki, Ibaraki 305-8687, Japan  
Tel: 81-298-73-3211(ex.444)

email: uaiuoc@ipri.aifrc.go.jp;  
URL: <http://www.fipri.aifrc.go.jp/labs/cjgenome/database/cjdatae.htm>  
1.

```

SOURCE
1. 490
LOCATION/YUAILLETS
/organism="Cryptomeria japonica"
/db_xref="taxon:3369"

```

BASE COUNT	135 a	99 c	141 g	215 t
ORIGIN				

Query Match	39.5%	Score 48.6;	DB 10;	Length 490;
Best Local Similarity	80.3%	Pred. No. 1.7e-05;		
Matches 57; Conservative	0;	Mismatches 14;	Totals 0;	Cases 0;

[illegible]

QY	64	cactgactgt	74
Dh	478	TTAATCACTGCT	488

RESULT	41
BG606610	
LOCUS	558 bp mRNA
DEFINITION	WHE2958.F12.L2425 Wheat dormant embryo cDNA library Triticum aestivum cDNA clone WHE2958.F12.L24, mRNA sequence.
ACCESSION	BG606610

```

VERSION      BG606610.1  GI:13656593
RECORDS      1
PAGE

```

SOURCE	bread wheat.
ORGANISM	<i>Triticum aestivum</i>

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophytes;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
; Triticeae; Triticum.  
; Triticum; FEA

TITLE	Author
The structure and function of the expressed portion of the wheat genomes - Dormant embryo cDNA library	Rausch, G.R., Walker-Simmons, M.K. and Wilson, C., D., Lazo, G.R.

**COMMENT**  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center

Tel: 5105595773  
Fax: 5105595818  
Email: [caandersn@pw.usda.gov](mailto:caandersn@pw.usda.gov)

quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.  
Location/Qualifiers

```

/organism="Triticum aestivum"
/cultivar="Brevor"
/db_xref="taxon:4565"

```

```
/clone_lib="Wheat dormant embryo cDNA library"
/tissue_type="Seed embryo"
/dev_stage="Mature seed"
```

/note="vector": lambda uni Zap AR, excised pnaagemid;  
Site\_1: ECORI; Site\_2: XhoI; Plants were grown to seed  
maturity under conditions favoring seed dormancy (L.  
Doherty at K. Walker-Simmons lab, Washington State

BASE COUNT	162 a	115 c	149 g	131 t	1 others
ORIGIN	Anderson lab	(all other authors).	"	preparations and DNA sequencing were performed by	

Query Match	39.5%;	Score 48.6;	DB 11;	Length 558;
Best Local Similarity	80.3%;	Pred. No. 1.8e-05;		
Matches 57; Conservative	0;	Mismatches 14;	Indels 0;	Gaps

QY	4	gtattgtcaaatgtga
pb	352	ggttttggtaggtgtg

QY	64	cactgactggt	74
Db	412	CTCTGATTGGT	422



















PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150984.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## alignment\_scores:

Ratio: 138.00 Length: 23  
Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 91.304

## alignment\_block:

US-09-684-016-48411 x AG05461 ..

Align seg 1/1 to: AG05461 from: 1 to: 186







PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149175.  
PR 20-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149722.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 27-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores: 138.00 Length: 23  
Ratio: 6.273 Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 91.304  
alignment\_block:

US-09-684-016-48411 x AAC06765 ..  
Align seq 1/1 to: AAC06765 from: 1 to: 186  
5 TATGTAAATGTGATGCCTTCAACCTGATGACCTATGCGGAATG 54  
|||||  
110 Tyrcyslucyscysglumetprotyrasnproaspaspmetvalglncy 126  
55 CGAGGCTGCACCTGACTGG 73  
|||||  
126 sGLUGlyCysLysaspTrp 132  
seq\_name: /SIDS2/gcgcdata/geneseq/geneseq/AA2000.DAT:AAG36118  
seq\_documentation\_block:  
ID AAG36118 standard; Protein; 186 AA.  
XX  
AC AAG36118;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 44217.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
XX Zea mays subsp. mays.  
XX  
XX EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137522.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.



PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139452.  
PR 17-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141642.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144684.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145215.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145921.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.

PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149930.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0152363.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154479.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 26-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158322.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.  
PR 22-OCT-1999; 9905-0160989.  
PR 25-OCT-1999; 9905-0161404.  
PR 25-OCT-1999; 9905-0161405.  
PR 25-OCT-1999; 9905-0161406.  
PR 26-OCT-1999; 9905-0161359.  
PR 26-OCT-1999; 9905-0161360.  
PR 26-OCT-1999; 9905-0161361.  
PR 28-OCT-1999; 9905-0161920.  
PR 28-OCT-1999; 9905-0161992.  
PR 28-OCT-1999; 9905-0161993.  
PR 29-OCT-1999; 9905-0162142.

alignment\_scores:

Quality: 138.00  
Ratio: 6.273Length: 23  
Gaps: 0



Percent Similarity: 95.652 Percent Identity: 91.304

alignment\_block:

US-09-684-016-48411 x AAG36118 ..

Align seg 1/1 to: AAG36118 from: 1 to: 186

```
5 TATGTGAATGTGATGCTTACACCCGTGATGACCTAATGGCGCAATG 54
|||||
110 TYRCYLSYCSYSLumetProTYRAsnProAspAspleuMetValGlnCY 126
55 CGAGGCTGCACCTGACTGCG 73
|||||
126 sGIuGLYCYSLyAspPTrP 132
```

seq\_name: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT:AAG39943

seq\_documentation\_block:

ID AAG39943 standard; Protein: 186 AA.

AC AAG39943;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49493.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-012180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 23-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.



PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158332.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:                    138.00                    Length:                    23  
  Quality:                    6.273                    Gaps:                    0  
Percent Similarity:                    95.652                    Percent Identity:                    91.304  
  
alignment\_block:  
us-09-684-016-48411 x AAG39943                    ..  
  
Align seg 1/1 to: AAG39943 from: 1 to: 186  
  
5 TATTGTAATGTGAGATGCTTACACCCGTGATGACCTAATGCGCAATG 54  
|||||  
110 TycrCysLysCysGluMetProTyrAsnProAspAspLeuMetValGlnCys 126  
55 CGAGGCTGCACCTGACTGG 73  
|||||  
126 sGluGlyCysLysAspTrp 132  
  
seq\_name: /SID2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG39942  
  
seq\_documentation\_block:  
ID AAG39942 standard; Protein; 191 AA.  
XX AAG39942;  
AC  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49492.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134270.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.



PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139751.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142053.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145912.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160816.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.



PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## alignment\_scores:

Quality: 138.00 Length: 23  
Ratio: 6.273 Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 91.304

## alignment\_block:

US-09-684-016-48411 x AAG39942 ..

Align seg 1/1 to: AAG39942 from: 1 to: 191

5 TATTGTAATGTGAGATGCCCTTACAACCTGATGACCTAATGGCGAATG 54  
|||||  
115 TrrCysLysCysGluMetProTyrAsnProAspAspLeuMetValGlnCys 131  
55 CGAGGGCTGCTACTGACTGG 73  
|||||  
131 scLugLysCysLysAspTrp 137

seq\_name: /SIDS2/gcdata/geneseq/geneseq/AA2000.DAT:AAG31123

seq\_documentation\_block:

ID AAG31123 standard; Protein: 196 AA.

AC AAG31123;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 37324.

KW Protein identification; signal transduction pathway; metabolic pathway;

KM hybridisation assay; genetic mapping; gene expression control; promoter;

OS Arabidopsis thaliana.

PN EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-01231825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 16-APR-1999; 99US-0128714.

XX 19-APR-1999; 99US-0129845.

XX 21-APR-1999; 99US-0130077.

XX 23-APR-1999; 99US-0130449.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 07-MAY-1999; 99US-0132486.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139464.

PR 18-JUN-1999; 99US-0139465.

PR 18-JUN-1999; 99US-0139466.

PR 18-JUN-1999; 99US-0139467.

PR 18-JUN-1999; 99US-0139468.

PR 18-JUN-1999; 99US-0139469.

PR 18-JUN-1999; 99US-0139470.

PR 18-JUN-1999; 99US-0139471.

PR 18-JUN-1999; 99US-0139472.

PR 18-JUN-1999; 99US-0139473.

PR 18-JUN-1999; 99US-0139474.

PR 18-JUN-1999; 99US-0139475.

PR 18-JUN-1999; 99US-0139476.

PR 18-JUN-1999; 99US-0139477.

PR 18-JUN-1999; 99US-0139478.

PR 18-JUN-1999; 99US-0139479.

PR 18-JUN-1999; 99US-0139480.

PR 18-JUN-1999; 99US-0139481.

PR 18-JUN-1999; 99US-0139482.

PR 18-JUN-1999; 99US-0139483.

PR 18-JUN-1999; 99US-0139484.

PR 18-JUN-1999; 99US-0139485.

PR 18-JUN-1999; 99US-0139486.

PR 18-JUN-1999; 99US-0139487.

PR 18-JUN-1999; 99US-0139488.

PR 18-JUN-1999; 99US-0139489.

PR 18-JUN-1999; 99US-0139490.

PR 18-JUN-1999; 99US-0139491.

PR 18-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139493.

PR 18-JUN-1999; 99US-0139494.

PR 18-JUN-1999; 99US-0139495.

PR 18-JUN-1999; 99US-0139496.

PR 18-JUN-1999; 99US-0139497.

PR 18-JUN-1999; 99US-0139498.

PR 18-JUN-1999; 99US-0139499.

PR 18-JUN-1999; 99US-0139500.

PR 18-JUN-1999; 99US-0139501.

PR 18-JUN-1999; 99US-0139502.

PR 18-JUN-1999; 99US-0139503.

PR 18-JUN-1999; 99US-0139504.

PR 18-JUN-1999; 99US-0139505.

PR 18-JUN-1999; 99US-0139506.

PR 18-JUN-1999; 99US-0139507.

PR 18-JUN-1999; 99US-0139508.

PR 18-JUN-1999; 99US-0139509.

PR 18-JUN-1999; 99US-0139510.

PR 18-JUN-1999; 99US-0139511.

PR 18-JUN-1999; 99US-0139512.

PR 18-JUN-1999; 99US-0139513.

PR 18-JUN-1999; 99US-0139514.

PR 18-JUN-1999; 99US-0139515.

PR 18-JUN-1999; 99US-0139516.

PR 18-JUN-1999; 99US-0139517.

PR 18-JUN-1999; 99US-0139518.

PR 18-JUN-1999; 99US-0139519.

PR 18-JUN-1999; 99US-0139520.

PR 18-JUN-1999; 99US-0139521.

PR 18-JUN-1999; 99US-0139522.

PR 18-JUN-1999; 99US-0139523.

PR 18-JUN-1999; 99US-0139524.

PR 18-JUN-1999; 99US-0139525.

PR 18-JUN-1999; 99US-0139526.

PR 18-JUN-1999; 99US-0139527.

PR 18-JUN-1999; 99US-0139528.

PR 18-JUN-1999; 99US-0139529.

PR 18-JUN-1999; 99US-0139530.

PR 18-JUN-1999; 99US-0139531.

PR 18-JUN-1999; 99US-0139532.

PR 18-JUN-1999; 99US-0139533.

PR 18-JUN-1999; 99US-0139534.

PR 18-JUN-1999; 99US-0139535.

PR 18-JUN-1999; 99US-0139536.

PR 18-JUN-1999; 99US-0139537.

PR 18-JUN-1999; 99US-0139538.

PR 18-JUN-1999; 99US-0139539.

PR 18-JUN-1999; 99US-0139540.

PR 18-JUN-1999; 99US-0139541.

PR 18-JUN-1999; 99US-0139542.

PR 18-JUN-1999; 99US-0139543.

PR 18-JUN-1999; 99US-0139544.

PR 18-JUN-1999; 99US-0139545.

PR 18-JUN-1999; 99US-0139546.

PR 18-JUN-1999; 99US-0139547.

PR 18-JUN-1999; 99US-0139548.

PR 18-JUN-1999; 99US-0139549.

PR 18-JUN-1999; 99US-0139550.

PR 18-JUN-1999; 99US-0139551.

PR 18-JUN-1999; 99US-0139552.

PR 18-JUN-1999; 99US-0139553.

PR 18-JUN-1999; 99US-0139554.

PR 18-JUN-1999; 99US-0139555.

PR 18-JUN-1999; 99US-0139556.

PR 18-JUN-1999; 99US-0139557.

PR 18-JUN-1999; 99US-0139558.

PR 18-JUN-1999; 99US-0139559.

PR 18-JUN-1999; 99US-0139560.

PR 18-JUN-1999; 99US-0139561.

PR 18-JUN-1999; 99US-0139562.

PR 18-JUN-1999; 99US-0139563.

PR 18-JUN-1999; 99US-0139564.

PR 18-JUN-1999; 99US-0139565.

PR 18-JUN-1999; 99US-0139566.

PR 18-JUN-1999; 99US-0139567.

PR 18-JUN-1999; 99US-0139568.

PR 18-JUN-1999; 99US-0139569.

PR 18-JUN-1999; 99US-0139570.

PR 18-JUN-1999; 99US-0139571.

PR 18-JUN-1999; 99US-0139572.

PR 18-JUN-1999; 99US-0139573.

PR 18-JUN-1999; 99US-0139574.

PR 18-JUN-1999; 99US-0139575.

PR 18-JUN-1999; 99US-0139576.

PR 18-JUN-1999; 99US-0139577.

PR 18-JUN-1999; 99US-0139578.

PR 18-JUN-1999; 99US-0139579.

PR 18-JUN-1999; 99US-0139580.

PR 18-JUN-1999; 99US-0139581.

PR 18-JUN-1999; 99US-0139582.

PR 18-JUN-1999; 99US-0139583.

PR 18-JUN-1999; 99US-0139584.

PR 18-JUN-1999; 99US-0139585.

PR 18-JUN-1999; 99US-0139586.

PR 18-JUN-1999; 99US-0139587.

PR 18-JUN-1999; 99US-0139588.

PR 18-JUN-1999; 99US-0139589.

PR 18-JUN-1999; 99US-0139590.

PR 18-JUN-1999; 99US-0139591.

PR 18-JUN-1999; 99US-0139592.

PR 18-JUN-1999; 99US-0139593.

PR 18-JUN-1999; 99US-0139594.

PR 18-JUN-1999; 99US-0139595.

PR 18-JUN-1999; 99US-0139596.

PR 18-JUN-1999; 99US-0139597.

PR 18-JUN-1999; 99US-0139598.

PR 18-JUN-1999; 99US-0139599.

PR 18-JUN-1999; 99US-0139600.

PR 18-JUN-1999; 99US-0139601.

PR 18-JUN-1999; 99US-0139602.

PR 18-JUN-1999; 99US-0139603.

PR 18-JUN-1999; 99US-0139604.

PR 18-JUN-1999; 99US-0139605.

PR 18-JUN-1999; 99US-0139606.

PR 18-JUN-1999; 99US-0139607.

PR 18-JUN-1999; 99US-0139608.

PR 18-JUN-1999; 99US-0139609.

PR 18-JUN-1999; 99US-0139610.

PR 18-JUN-1999; 99US-0139611.

PR 18-JUN-1999; 99US-0139612.

PR 18-JUN-1999; 99US-0139613.

PR 18-JUN-1999; 99US-0139614.

PR 18-JUN-1999; 99US-0139615.

PR 18-JUN-1999; 99US-0139616.

PR 18-JUN-1999; 99US-0139617.

PR 18-JUN-1999; 99US-0139618.



```

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

```

```

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 138.00      Length: 23
    Ratio: 6.273        Gaps: 0
    Percent Similarity: 95.652    Percent Identity: 91.304

alignment block:
US-09-684-016-48411 x AAG31123  ..

Align seg 1/1 to: AAG31123 from: 1 to: 196

5 TATGTAAATGTCGATGCCTTACACCGATGACCTAATGCGCAATG 54
|||||
110 TyrcyslyscysglumetPrtotyranPronspaspleumetvalGlnCy 126
55 CGAGGCTGCACCTGACTGG 73
|||||
126 sGLuGLyCysLysAspTrp 132

seq_name: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AAG31122
seq_documentation_block:
ID AAG31122 standard; Protein; 201 AA.
XX
AC AAG31122;
XX
DT 17-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37323.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.

```



PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138867.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139453.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140393.
PR 23-JUN-1999;	99US-0140394.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141642.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143624.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144684.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 23-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149375.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149930.
PR 23-AUG-1999;	99US-0150566.
PR 25-AUG-1999;	99US-0150884.
PR 26-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 18-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.
PR 22-OCT-1999;	99US-0160989.
PR 25-OCT-1999;	99US-0161404.
PR 25-OCT-1999;	99US-0161405.



PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
alignment\_scores:  
Quality: 138.00 Length: 23  
Ratio: 6.273 Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 91.304  
  
alignment\_block:  
US-09-684-016-48411 x AAG31122 ..  
  
Align seg 1/1 to: AAG31122 from: 1 to: 201  
  
5 TATTGTAATGTGAGATGCCTTACACCCGTGATGACCTAATGGCGCAATG 54  
|||||  
115 TyrcysLysCysSglumetProTyAsnProAspAspleuValGlnCy 131  
55 CGAGGCGTGCACGTGACTGG 73  
|||||  
131 sglumGlyCysLysAspTrp 137  
  
seq\_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG36117  
  
seq\_documentation\_block:  
ID AAG36117 standard; Protein; 216 AA.  
XX AAG36117;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 44216.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
XX Zea mays subsp. mays.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145193.  
PR 23-JUL-1999; 99US-0145218.



```

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147208.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.

```

```

PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 138.00
    Ratio: 6.273
    Percent Similarity: 95.652
    Percent Identity: 91.304

alignment_block:
US-09-684-016-48411 x AAG36117 ..

Align seg 1/1 to: AAG36117 from: 1 to: 216

5 TATTGTAATGTGAGATGCTTACACCCCTGATGACCTAATGCCCAATG 54
140 TTYCysLysCysGluMetProTyrAsnProAspAspLeuMetValGlnCy 156
55 CGAGGCTGCTGACTGCTGG 73
156 sGlnGlyCysLysAspTrp 162

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT: AAG05460
seq_documentation_block:
ID AAG05460 standard; Protein: 224 AA.
XX
AC AAG05460;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1874.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.

```



PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145513.  
PR 27-JUL-1999; 99US-0145818.  
PR 27-JUL-1999; 99US-0145819.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156556.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157733.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159299.  
PR 13-OCT-1999; 99US-0159325.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.



PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161820.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:                   Quality: 138.00                   Length: 23  
                                  Ratio: 6.273                   Gaps: 0  
Percent Similarity: 95.652       Percent Identity: 91.304

alignment\_block:

US-09-684-016-48411 x AAG05460 ..

Align seg 1/1 to: AAG05460 from: 1 to: 224

5 TATTGTAATGTGAGATGCTTACACCCGTGATGACCTATATGCGCATG 54  
|||||  
148 TYTCYSLSYSCYSGLUMETPROTYRASPROMSPLEUMETVALGNCY 164  
55 CGAGGGCTGCACCTGACTGG 73  
|||||  
164 SGLUGLYCYSLYSASPTRP 170

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAG36116

seq\_documentation\_block:

ID AAG36116 standard; Protein; 338 AA.

XX AAG36116;

XX 18-OCT-2000 (first entry)

XX Zea mays protein fragment SEQ ID NO: 44215.

KM Protein identification: signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KM termination sequence; corn.

XX Zea mays subsp. mays.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-01231825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.



```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145386.
PR 02-AUG-1999; 99US-0145388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156566.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

```

```

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 138.00      Length: 23
    Ratio: 6.273        Gaps: 0
Percent Similarity: 95.652  Percent Identity: 91.304

alignment_block:
US-09-684-016-48411 x AAG36116 ..

Align seg 1/1 to: AAG36116 from: 1 to: 338

5 TATGTAAATGTCGATGCCTTCAACCGATGACCTATATGCGCAATG 54
|||||
262 TyrcyslscysglumetProtyrAsnProaspSplumetValGncy 278
55 CGAGGCTGCACCTGACTGTG 73
|||||
278 sGluGlyCysLysAspTrr 284

seq_name: /STDs2/gcgcdata/geneseq/geneseqp/AA2000.DAT:AAG04540
seq_documentation_block:
ID AAG04540 standard; Protein; 198 AA.
XX
AC AAG04540;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 604.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.

```



PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148584.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.



PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## alignment\_scores:

Quality: 124.00 Length: 23  
Ratio: 5.636 Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 78.261

## alignment\_block:

US-09-684-016-48411 x AAG04540 ..

Align seg 1/1 to: AAG04540 from: 1 to: 198

5 TATGTAAATGATGATGCTTACACCCGTGATGACCTAATGGCGCAATG 54

111 PheCysLysCysGlnMetProTyrAsnProAspAspMetValGlnCys 127

55 CGAGGCGCTGCAGTACTGG 73

127 sGlnGlnCysSerGlnTyr 133

seq\_name: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT: AAG38175

seq\_documentation\_block:

ID AAG38175 standard; Protein: 198 AA.

AC AAG38175;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 47057.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126284.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 16-APR-1999; 99US-0128714.

XX 19-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139494.  
PR 18-JUN-1999; 99US-0139495.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.



PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148655.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149222.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0153486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158332.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0158295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160778.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## alignment\_scores:

Quality:	124.00	Length:	23
Ratio:	5.636 <td>Gaps:</td> <td>0</td>	Gaps:	0
Percent Similarity:	95.652	Percent Identity:	78.261

## alignment\_block:

US-09-684-016-48411 x AAG38175 ..

Align seg 1/1 to: AAG38175 from: 1 to: 198

5 TATTGTAATGTGAGATGCCCTTACACCCGTGATGACCTAATGGCGCAATG 54

111 PheCysIysCysGluMetProTyrAsnProAspAspIleuMetValGlnCys 127

55 CGAGGCGTGCACACTGCTGG 73

127 sGIuGlucYsSerGIuTrp 133

seq\_name: /SID22/gcgdata/geneseq/geneseqp/AA2000.DAT: AAG04539

seq\_documentation\_block:

ID AAG04539 standard; Protein; 228 AA.

AC AAG04539;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 603.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.



PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138099.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147203.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.



```
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
alignment_scores:
  Quality: 124.00      Length: 23
  Ratio: 5.636        Gaps: 0
Percent Similarity: 95.652  Percent Identity: 78.261
```

alignment\_block:

US-09-684-016-48411 x AAG04539 ..

Align seg 1/1 to: AAG04539 from: 1 to: 228

```
5 TATGTAAATGTGAGATGCTTACAAACCGTGATGACCTAATGGCGCAATG 54
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
141 PhcYslYsCysGluMetProTyrAsnProAspAspLeuMetValGlnCy 157
55 CGAGGGCTGCACCTGACTGG 73
  ||| ||:::||||
157 sgluGluCysSerGluTrp 163
```

seq\_name: /SITS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG38174

seq\_documentation\_block:

ID AAG38174 standard; Protein; 228 AA.

```
XX
AC AAG38174;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47056.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
```

```
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132853.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140635.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
```



```

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154038.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

```

```

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

alignment_scores:      Quality: 124.00      Length: 23
                        Ratio: 5.636        Gaps: 0
Percent Similarity: 95.652  Percent Identity: 78.261

```

```

alignment block:
US-09-684-016-48411 x MAG38174 ..

```

```

Align seq 1/1 to: MAG38174 from: 1 to: 228

```

```

5 TATGTGAATGTGATGCTTACAACTGATGACTAATGCGCATG 54
:::|||||
141 PhcCysLysCysGslUmMetProTyrAsnProAspSplumetValIndcY 157
55 CGAGGCGCTGCACGTGACTGG 73
|||||
157 sglUglucYserGluTrp 163

```

```

seq_name: /stds/gcgdata/geneseq/geneseq/AA2000.DAT: MAG40195

```

```

seq_documentation_block:

```

```

ID MAG40195 standard; Protein; 270 AA.

```

```

AC MAG40195;

```

```

DT 18-OCT-2000 (first entry)

```

```

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49840.

```

```

KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

```

```

KW Arabidopsis thaliana.

```

```

OS EPI033405-A2.

```

```

PN 06-SEP-2000.

```

```

PD 25-FEB-2000; 2000EP-0301439.

```

```

PF 25-FEB-1999; 99US-0121825.

```

```

XX 05-MAR-1999; 99US-0123180.

```



PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135153.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139457.  
PR 16-JUN-1999; 99US-0139457.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141642.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.



PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:                   Quality: 124.00                   Length: 23  
                                  Ratio: 5.636                   Gaps: 0  
Percent Similarity: 95.652       Percent Identity: 78.261

alignment\_block:

US-09-684-016-48411 x AAG40195   ..

Align seg 1/1 to: AAG40195 from: 1 to: 270

5 TATGTGAATGTGATGCGCTTACACCCCTGATGACCTAATGGCGCATG 54  
                                  :::|||||                                  :::|||||  
183 PheCysLysCysGluMetProTyrAsnProAspAspLeuMetValGlnCys 199  
55 CGAGGCGTCGACTGACTGG 73  
      ||||                  ||||  
199 sgluGluCysSerGluTrrp 205

seq\_name: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG40194

seq\_documentation\_block:

ID AAG40194 standard; Protein: 357 AA.

XX AAG40194;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49839.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.



PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## alignment\_scores:

Quality: 124.00 Length: 23  
Ratio: 5.636 Gaps: 0  
Percent Similarity: 95.652 Percent Identity: 78.261

## alignment\_block:

US-09-684-016-48411 x AAG40194 ..

Align seq 1/1 to: AAG40194 from: 1 to: 357

5 TATTGTAATGAGATGCTTACACCCCTGATGACCTAATGCGCAATG 54  
:::|||||  
270 PhcYslYcysGluMetProTyrAsnProspaspseuMetValGlnCy 286

55 CGAGGCTGCACCTGCTGG 73  
|||||  
286 sgluGluCysserGluTrrp 292

seq\_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT.AAG40193

seq\_documentation\_block:

ID AAG40193 standard; Protein; 367 AA.

AC AAG40193;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49838.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

XX PN EP1033405-A2.



PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 05-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148688.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151308.  
PR 30-AUG-1999; 99US-0151308.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.



PR	23-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0155458
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157665
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159331
PR	14-OCT-1999	9905-0159637
PR	14-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159884
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160989
PR	22-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161405
PR	26-OCT-1999	9905-0161356
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	26-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161993
PR	28-OCT-1999	9905-0161993
PR	29-OCT-1999	9905-0162142

```
alignment_scores:
  Quality: 124.00
  Ratio: 5.636
  Percent Similarity: 95.652
  Length: 23
  Gaps: 0
  Percent Identity: 78.261
```

alignment\_block:

US-09-684-016-48411 X AAG40193 ..

Align seg 1/1 to: AAG40193 from: 1 to: 367

5 TATTGTAATATGTASATGTGCTTACAACCCGTATGACCTAAAGGGCCATG 54  
 ...|||||  
 280 PhcYalysCysLumetProTyrAsnProAspSpleUmetValGlnCys 296  
 55 CGAGGGCTGCATCTGACTCG 73  
 ...|||||  
 296 sGluGlnCysSerGluTrrp 302

seq\_name: /SIDS2/gcdata/geneseq/geneseq/A2001.DAT: AAB93780

seq\_documentation\_block:

ID	Protein; 790 AA.
AAB93780	standard;

AC AAB93780;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:13511.

**KW** Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

XX	07-FEB-2001.
PD	
XX	
PF	28-JUL-2000; 2000EP-0116126.

PA (HELI-) HELIX RES INST

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Prime sets for synthesizing polynucleotides, particularly the 5602 PT full-length cDNAs defined in the specification, and for the detection PT and/or diagnosis of the abnormality of the proteins encoded by the PT full-length cDNAs -

PS Claim 8; SEQ ID 13511; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA9892446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

**SQ Sequence 790 AA;**

[illegible]

```
alignment_block:
```

US-09-684-016-48411 X AAB93780 .

Align seg 1/1 to: AAB93780 from: 1 to: 790

[illegible]

seq\_name: /SIDS2/gcdata/geneseq/geneseqp/AA2000.DAT: AAB42371

seq\_documentation\_block:

ID AAB42371 standard; Protein; 1084 AA.



XX AAB42371;  
 AC  
 XX 08-FEB-2001 (first entry)  
 DT  
 XX  
 DE Human ORFX ORF2135 polypeptide sequence SEQ ID NO:4270.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnereary; antiparietal; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;  
 KW antidiabetic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;  
 KW thrombosis; contraceptive.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO2000058473-A2.  
 PD  
 XX 05-OCT-2000.  
 XX  
 PE 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX  
 PI Shimkets RA, Leach M;  
 XX  
 DR WPI: 2000-602362/57.  
 DR N-PSDB; AAC76580.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 3476-3478; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnereary;  
 CC antiparietal; antiparkinsonian; neurotropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineoplastic; antibacterial; antiviral; antifungal; antineoplastic;  
 CC antihypertensive; and antidiabetic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, diabetes mellitus,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC  
 XX Sequence 1084 AA;

alignment\_scores:      74.00      Length:      23  
                          quality:      4.625      Gaps:      0  
                          Ratio:      69.565      Percent Identity: 43.478

alignment\_block:  
 US-09-684-016-48411 x AAB42371 ..

Align seq 1/1 to: AAB42371 from: 1 to: 1084

5 TATGTAAAGTGGATGCTTACACCGCTGAGACCTAATGCGCAATG 54  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 67 Tyrosylleucylsargineprotyralspvaltharphenethylleucyl 83  
 55 CGAGCGCTGCACCTGACTGCG 73  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 83 saspmeccysglinsprrp 89

seq\_name: /STD2/gcgdata/geneseq/geneseq/AA2001.DAT:AA40982

seq\_documentation\_block:  
 ID AAM40982 standard; Protein: 384 AA.

AC AAM40982;  
 XX  
 XX

DT 22-OCT-2001 (first entry)  
 XX  
 XX

DE Human polypeptide SEQ ID NO 5913.  
 XX  
 XX

KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 KW  
 XX

OS Homo sapiens.  
 XX  
 XX

PN WO200153312-A1.  
 XX  
 XX

PD 26-JUL-2001.  
 XX  
 XX

PE 26-DEC-2000; 2000WO-US34263.  
 XX  
 XX

PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 XX

PA (HYSE-) HYSEQ INC.  
 XX  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AA160138.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 2; SEQ ID NO 5913; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and  
 XX the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and



CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 384 AA;  
  
alignment\_scores:  
Quality: 66.50 Length: 29  
Ratio: 4.433 Gaps: 2  
Percent Similarity: 51.724 Percent Identity: 48.276  
  
alignment\_block:  
US-09-684-016-48411 x AAM40982 ..  
  
Align seg 1/1 to: AAM40982 from: 1 to: 384  
  
5 TATTGTAATGTGAGATGCTTACCAACCTGAT..... 37  
||||| |||::: ||||| |||||  
137 TyrCysIleCysLysArgProTyr...ProAspProGluAspGluIlePr 152  
  
38 .GACCTAATGGCGCAATGCGAGGCTGCACCTGACTGG 73  
||| ||| ||||| ||| |||||  
152 oAspGluMetIleGlnCysValIValCysGluAspTrp 164  
  
seq\_name: /SIDS2/gcgdata/geneseq/geneseq/AAM2001.DAT:AAM39196  
  
seq\_documentation\_block:  
ID AAM39196 standard; Protein: 425 AA.  
XX  
AC AAM39196;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2341.  
XX  
KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO20015312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PE 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HXSE-) HXSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang Z, Wang J, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX

DR WPI: 2001-442253/47.  
DR N-PSDB; AA158352.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 4; SEQ ID NO 2341; 10078bp; English.  
XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 425 AA;  
  
alignment\_scores:  
Quality: 66.50 Length: 29  
Ratio: 4.433 Gaps: 2  
Percent Similarity: 51.724 Percent Identity: 48.276  
  
alignment\_block:  
US-09-684-016-48411 x AAM39196 ..  
  
Align seg 1/1 to: AAM39196 from: 1 to: 425  
  
5 TATTGTAATGTGAGATGCTTACCAACCTGAT..... 37  
||||| |||::: ||||| |||||  
134 TyrCysIleCysLysArgProTyr...ProAspProGluAspGluIlePr 149  
  
38 .GACCTAATGGCGCAATGCGAGGCTGCACCTGACTGG 73  
||| ||| ||||| ||| |||||  
149 oAspGluMetIleGlnCysValIValCysGluAspTrp 161  
  
seq\_name: /SIDS2/gcgdata/geneseq/geneseq/AAM2001.DAT:AAB60498  
  
seq\_documentation\_block:  
ID AAB60498 standard; Protein: 425 AA.  
XX  
AC AAB60498;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Human cell cycle and proliferation protein CCYPR-46, SEQ ID NO:46.  
XX  
KW Cell cycle and proliferation protein; CCYPR; human; agonist;  
KW antagonist; gene therapy; detection; gene therapy;  
KW transgenic animal disease model; immune disorder;  
KW developmental disorder; cell signalling disorder;  
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
KW menstrual cycle disorder; bacterial infection.  
XX  
OS Homo sapiens.  
XX  
PN WO200107471-A2.  
XX  
PD 01-FEB-2001.  
XX  
PE 21-JUL-2000; 2000WO-US19948.  
XX







PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145216.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149388.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149922.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161993.  
PR 28-OCT-1999; 99US-0162142.

alignment\_scores:  
Quality: 65.00 Length: 14  
Ratio: 5.417 Gaps: 0  
Percent Similarity: 85.714 Percent Identity: 78.571

alignment\_block:

US-09-684-016-48411 x AAG47973 ..

Align seg 1/1 to: AAG47973 from: 1 to: 193

5 TATTGTAATGTCAGATGCTTACAAACCTGTGACCTTAATG 46  
||||| ||||||| ||||||| ||||||| ||||||| |||||||  
110 TYRTRYLYSCYSGLUMETPROTYRASNRSERASPGIUMET 123

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/AAY04323.DAT:AAY04323

seq\_documentation\_block:

ID AAY04323 standard; Protein: 238 AA.

XX AAY04323;

DT 18-JUN-1999 (first entry)

XX Human secreted protein SEQ ID NO:69.

XX Human; secreted protein; cancer; tumour; developmental abnormality;



```

KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS.
XX
OS Homo sapiens.
XX
PM WO9910363-A1.
PD 04-MAR-1999.
XX
PE 27-AUG-1998; 98WO-US17709.
PR 29-AUG-1997; 97US-0056271.
PR 29-AUG-1997; 97US-0056073.
PR 29-AUG-1997; 97US-0056247.
PR 29-AUG-1997; 97US-0056270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF:
DR WPI: 1999-190585/16.
XX
PT New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Disclosure: Page 18; 170pp: English.
XX
CC AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.
CC CC AAU04293 to AAU04321 represent the secreted proteins encoded by the 29
CC human genes. The genes and their corresponding secreted polypeptides are
CC useful for preventing, treating or ameliorating medical conditions,
CC e.g. by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new genes. Specific
CC uses are described for each of the 29 genes, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC polypeptides are also useful for identifying their binding partners.
CC The sequences given in AAX30174 to AAX30182 and AAU04322 to AAU04334 are
CC used in the exemplification of the present invention.
XX
XX Sequence 238 AA:
XX
alignment_scores:
Quality: 60.00 Length: 23
Ratio: 4.000 Gaps: 0
Percent Similarity: 65.217 Percent Identity: 34.783
Alignment_block:
US-09-684-016-48411 x AAU04323 ..
Align seg 1/1 to: AAU04323 from: 1 to: 238
5 TATTGTAATGTGAGATGGCTTACACCCTGATGACTAATGGCGCAATG 54
||||| ||||| : ||||| : : ||||| : ||
61 TYTCYSILIECYSAIRGTHPRTFYRASPGLNSERGINPHETYLIEGLYCY 77
CGAGGCCTGCCTGACTGCTGG 73
|::: ||| ::|||
77 SASPARGYCSGINASHTTP 83

```

```
seq_documentation_block:
ID AAY57453 standard; Protein: 2781 AA.
XX
AC AAY57453;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human transcriptional regulatory factor SEQ ID NO:10.
XX
KW Human: transcriptional regulatory factor; TCOA1; BLAST detection; bromo-domain; cell proliferation; cancer.
XX
OS Homo sapiens.
XX
PN W0957143-A1.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-JP02340.
XX
PR 30-APR-1998; 98JP-0137631.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Jones MH;
XX
DR WPI; 2000-052940/04.
DR N-PSDB; AAZ39033.
XX
PT Transcriptional regulatory factor containing a bromo domain and gene TCOA1 encoding it -
XX
PS Claim 1; Page 139-151; 154pp; Japanese.
XX
CC The present sequence represents a human transcriptional regulatory factor containing a bromo domain. The factor interacts with proteins involved in the chromatin-mediated transcription regulation mechanism. It binds CC to hSNF2H, hSNF2L and NCOA-62/Skip. It can be used for screening compounds binding to it and acting as agonists or antagonists, which CC are potentially useful for the treatment and prevention of cancer and other cell proliferation disorders.
XX
SQ Sequence 2781 AA;

alignment_scores:
Quality: 60.00 Length: 23
Ratio: 4.286 Gaps: 0
Percent Similarity: 60.870 Percent Identity: 34.783

alignment_block:
US-09-684-016-48411 x AAY57453 ..

Align seq 1/1 to: AAY57453 from: 1 to: 2781

5 TATGTTAATGAGATGCCTTACACCGTAGCCTAAATGGCGCANG 54
||||| ||||| :|||::: ::::: ||
2604 TyrCysIleCysIylThrProTyAspGluserylsherylleGlycy 2620
55 CGAGGCGTGCACTGCTGG 73
1::: ||| ::|||
2620 sAspargCysglInsntRp 2626

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY57452

seq_documentation_block:
ID AAY57452 standard; Protein: 2907 AA.
XX
AC AAY57452;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human transcriptional regulatory factor SEQ ID NO:1.
```



XX	06-SEP-2000.	2000EP-0301439
PD		
XX		
PF	25-FEB-2000;	2000EP-0301439
PR	25-FEB-1999;	99US-01231825
PR	03-MAR-1999;	99US-01231825
PR	05-MAR-1999;	99US-01233548
PR	22-MAR-1999;	99US-01257598
PR	25-MAR-1999;	99US-01262654
PR	29-MAR-1999;	99US-0126785
PR	01-APR-1999;	99US-01272652
PR	04-APR-1999;	99US-01282534
PR	08-APR-1999;	99US-01287114
PR	16-APR-1999;	99US-01298445
PR	19-APR-1999;	99US-01300077
PR	21-APR-1999;	99US-01300449
PR	22-APR-1999;	99US-01305010
PR	23-APR-1999;	99US-01308691
PR	28-APR-1999;	99US-01314449
PR	APR-1999;	99US-01320048
PR	30-APR-1999;	99US-01324078
PR	05-MAY-1999;	99US-01324884
PR	05-MAY-1999;	99US-01332885
PR	06-MAY-1999;	99US-01332886
PR	07-MAY-1999;	99US-01328683
PR	11-MAY-1999;	99US-01342556
PR	14-MAY-1999;	99US-01343218
PR	14-MAY-1999;	99US-01343219
PR	14-MAY-1999;	99US-01343220
PR	14-MAY-1999;	99US-01343221
PR	18-MAY-1999;	99US-01347461
PR	19-MAY-1999;	99US-01353524
PR	20-MAY-1999;	99US-01355124
PR	21-MAY-1999;	99US-01353553
PR	24-MAY-1999;	99US-01355629
PR	25-MAY-1999;	99US-01360621
PR	27-MAY-1999;	99US-01367882
PR	28-MAY-1999;	99US-01372722
PR	01-JUN-1999;	99US-01373528
PR	03-JUN-1999;	99US-01377502
PR	04-JUN-1999;	99US-01377524
PR	07-JUN-1999;	99US-01380540
PR	08-JUN-1999;	99US-01385840
PR	10-JUN-1999;	99US-01388447
PR	14-JUN-1999;	99US-01391119
PR	16-JUN-1999;	99US-01394512
PR	16-JUN-1999;	99US-01394653
PR	17-JUN-1999;	99US-01394654
PR	18-JUN-1999;	99US-01394654
PR	18-JUN-1999;	99US-01394655
PR	18-JUN-1999;	99US-01394656
PR	18-JUN-1999;	99US-01394657
PR	18-JUN-1999;	99US-01394658
PR	18-JUN-1999;	99US-01394659
PR	18-JUN-1999;	99US-01394660
PR	18-JUN-1999;	99US-01394661
PR	18-JUN-1999;	99US-01394662
PR	22-JUN-1999;	99US-01403554
PR	23-JUN-1999;	99US-01406055
PR	24-JUN-1999;	99US-01406055
PR	28-JUN-1999;	99US-01408023
PR	29-JUN-1999;	99US-01409181
PR	30-JUN-1999;	99US-0141287
PR	01-JUL-1999;	99US-01418442
PR	01-JUL-1999;	99US-01421541



PR	06-JUL-1999	9905-0142055
PR	02-JUL-1999	9905-0142059
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142820
PR	12-JUL-1999	9905-0142877
PR	13-JUL-1999	9905-0143342
PR	14-JUL-1999	9905-0143362
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144332
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144532
PR	20-JUL-1999	9905-0144584
PR	21-JUL-1999	9905-0144814
PR	21-JUL-1999	9905-0145086
PR	21-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145085
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145089
PR	22-JUL-1999	9905-0145192
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	26-JUL-1999	9905-0145324
PR	26-JUL-1999	9905-0145376
PR	27-JUL-1999	9905-0145813
PR	27-JUL-1999	9905-0145818
PR	27-JUL-1999	9905-0145819
PR	28-JUL-1999	9905-0145951
PR	02-AUG-1999	9905-0146386
PR	02-AUG-1999	9905-0146388
PR	02-AUG-1999	9905-0146589
PR	03-AUG-1999	9905-0147038
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147192
PR	05-AUG-1999	9905-0147260
PR	11-AUG-1999	9905-0148341
PR	13-AUG-1999	9905-0148565
PR	13-AUG-1999	9905-0148684
PR	16-AUG-1999	9905-0149368
PR	17-AUG-1999	9905-0149175
PR	18-AUG-1999	9905-0149426
PR	26-AUG-1999	9905-0149722
PR	20-AUG-1999	9905-0149723
PR	20-AUG-1999	9905-0149829
PR	23-AUG-1999	9905-0149802
PR	23-AUG-1999	9905-0149930
PR	26-AUG-1999	9905-0150566
PR	26-AUG-1999	9905-0150884
PR	27-AUG-1999	9905-0151065
PR	27-AUG-1999	9905-0151066
PR	30-AUG-1999	9905-0151303
PR	31-AUG-1999	9905-0151348
PR	01-SEP-1999	9905-0151930
PR	07-SEP-1999	9905-0152363
PR	13-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	16-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154779

PR	22-SEP-1999	9905-0155148
PR	23-SEP-1999	9905-0155549
PR	24-SEP-1999	9905-0155659
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156586
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157665
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	13-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159331
PR	14-OCT-1999	9905-0159637
PR	14-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160989
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161350
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161993
PR	28-OCT-1999	9905-0161992
PR	29-OCT-1999	9905-0162142

alignment_scores:		
quality:	57.00	length: 24
ratio:	3.167	gaps: 0
Percent Similarity:	75.000	Percent Identity: 33.333

alignment\_block:

US-09-684-016-48411/rev x AAG51002

Align seg 1/1 to: AAG51002 from: 1 to: 480

104 AAGTAAAGTATCCACCTTAACTGTGAACCTTACCATGACATGACAGGCTCG 55  
 |||:|||||: |||  
 161 TysleuansmleucutylentstleutyValTlntrArglnSerGluProcy 177  
 54 CATGGCCCCATTAGGTCATACG 33  
 :|||:|||||:  
 177 sleuGluProcyGlnHisGlu 184

seq\_name: /SIDS2/gcdata/geneseq/geneseqp/AA2000.DAT:AA651001

```
seq_documentation_block:
ID   AACG51001 standard; Protein; 567 AA.
```

AC AAG51001;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 64687.

KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX



```
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137232.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148645.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
```



PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:                   Quality: 57.00                   Length: 24  
                                  Ratio: 3.167                   Gaps: 0  
Percent Similarity: 75.000   Percent Identity: 33.333

alignment\_block:

US-09-684-016-48411/rev x AAG51001   ..

Align seg 1/1 to: AAG51001 from: 1 to: 567

104 AGATAGTATTCACCTTACATGTGAACCTTACGACGACGACGCTCG 55  
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
248 LysLeuAsnLeuGluIleHisIleTyValThrArgGlnSerGluProCy 264  
54 CATTGGCCGCAATTAGTCATCAG 33  
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
264 sleuGlyProTyGlyHisGlu 271

seq\_name: /SIDS2/cgcdata/geneseq/geneseq/AA2000.DAT:AAG51000

seq\_documentation\_block:

ID AAG51000 standard; Protein; 614 AA.

XX AAG51000;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 64686.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX Arabidopsis thaliana.  
OS EPI033405-A2.  
PN 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
PF XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.



```

PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

```

```

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

alignment_scores: 57.00      Length: 24
                   Ratio: 3.167      Gaps: 0
Percent Similarity: 75.000      Percent Identity: 33.333

```

alignment\_block:

```

US-09-684-016-48411/rev x AAG51000 ..
Align seg 1/1 to: AAG51000 from: 1 to: 614

```

```

104 AAGATNAGATACCTTACATGTGAACCTTACATCTACGACCCCTCG 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 LysIleuAsnIleuGlnIleHisIleTyValThrArgIleSerGluProCys 311

```

```

54 CATTCGCCCATTAGGTGATCAG 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 sIeuGlyProTyGlyGlnHisGlu 318

```

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AA043724

```

```

seq_documentation_block:

```

```

ID AAB43724 standard; Protein: 181 AA.

```

```

AC AAB43724;

```

```

XX 08-FEB-2001 (first entry)

```

```

DT
XX

```



```

DE Human cancer associated protein sequence SEQ ID NO:1169.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytosolic; proliferative; vulerary; immunomodulator;
KW antidiabetic; antiaschmatic; antirheumatic; antiarthritis; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PN 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WP1: 2000-587533/55.
XX
DR N-PSDB; AAC77933.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 11; Page 1790-1791; 2352pp; English.
XX
AC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB83338 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiaschmatic; antirheumatic; antiarthritis;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 181 AA;
XX
alignment_scores:
XX Quality: 56.50 Length: 23
XX Ratio: 3.531 Gaps: .1
XX Percent Similarity: 69.565 Percent Identity: 39.130
XX
alignment_block:
XX US-09-684-016-48411 x AAB43724 ..
XX
Align seg 1/1 to: AAB43724 from: 1 to: 181
XX
5 TATTGTGAATGTGAGATGCCTTACAACCGTAGACGACCTAATGGCGCAATG 54
XX ||||| |||:|||||:||||| :|| :|||
XX 26 TTYCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 41

```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA67579
seq_documentation_block:
ID   AA67579 standard; Protein; 562 AA.
XX                                     |::: ||| ::||| 
AC    AAY67579;                       41 sasparGcsgIugIntTp 47
DT      19-JUN-2000 (first entry)
DE    Human death inducer-obliterator 1 (DIO-1) polypeptide.
XX
KM     Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; human;
KW     autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW     antineumatic; antiinflammatory; antiproliferative.
XX
OS    Homo sapiens.
FH      Key              Location/Qualifiers
FT       Region          165..172
                          /note= "MLS sequence"
FT       Region          185..193
                          /note= "MLS sequence"
FT       Region          271..288
                          /note= "zinc finger motif"
FT       Region          293..320
                          /note= "zinc finger motif"
XX
PN     WO200015787-A1.
PD
PI     23-MAR-2000.
PE
PF      10-SEP-1999;    99WO-GB03019.
PX
PR      10-SEP-1998;    98SE-0003069.
PP      17-SEP-1998;    98US-0100873.
XX
XX (CNLS ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA (BANNN/) BANNERMAN D G.
PY
ALonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
DR WPI: 2000-271426/23.
DR N-PDB; AAZ90578.
XX
XX New DNA encoding human and murine death inducer-obliterator 1
PT polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT hyperproliferative skin disorders -
XX
Claim 6; Fig 1C; 27pp; English.
PS
XX The invention provides nucleic acids encoding the human and murine death
CC inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC agonists and antagonists are used as a medicament for treating diseases
CC characterized by an alteration in cell death or by hyperproliferation,
CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC tumours, malignant tumours or hyperproliferative skin disorders. They are
CC also useful in the treatment of metabolic, proliferative or inflammatory
CC conditions. The present sequence represents the human DIO-1 polypeptide.
CX
SQ Sequence 562 AA;
```











```
XX Hayden MR, Kalchman M;
XX
XX WPI: 1997-297870/27.
DR N-PSDB; AAT67187.
XX
XX cDNA encoding Huntington's disease interacting protein - useful to
PT ameliorate effects of disease in patient expressing Huntingtin
PT protein with expanded CAG repeat region
XX
PS Claim 4; Page 28-31; 39pp: English.
XX
XX Huntingtin interacting protein HIP1 (AAW18030) is a brain-specific
CC protein that interacts differently with the gene product of a
CC normal (16 CAG repeat) and an expanded (over 44 CAG repeat)
CC Huntington's disease (HD) gene, and which facilitates incorporation
CC of huntingtin into brain cell membranes. Its amino acid sequence
CC was deduced from a human frontal cortex cDNA clone (AAT67187). The
CC effects of HD in a patient expressing huntingtin protein with an
CC expanded polyglutamine tract can be ameliorated by increasing the
CC amount of expressed HIP1 in the brain using gene therapy
CC approaches. Modified forms of HIP1 which bind more effectively to
CC expanded HD protein can be used to convert the expanded protein
CC into a functional molecule.
XX
SQ Sequence 914 AA:

alignment_scores:
Quality: 54.00 Length: 38
Ratio: 2.077 Gaps: 1
Percent Similarity: 68.421 Percent Identity: 36.842

alignment_block:
US-09-684-016-48411/rev x AAW18030 ..
Align seg 1/1 to: AAW18030 from: 1 to: 914

119 GTGTGAATAAATGAGATAGATACCTTACATGGAACCTTACCAG 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 ValPheAsnSerLeuAspMetSerArgSerValSerValThrAlaIaIaG1 95
69 TCAGTGCAGCCCTGCATTCGCCCATTAAGTCATCAGGGTTGTAAGCAT 20
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 yGlnCys.....ArgLeuAlaProLeuIleGlnValIleLeuAspCys 110
19 CTCACATTTTACAT 6
|||||:|||||:|||||:|||||:|||||:|||||:
110 erHisLeuTyTAsp 114

seq_name: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT.AAY59269
seq_documentation_block:
ID AAY59269 standard; Protein; 914 AA.
XX
XX AAY59269;
XX
DT 17-APR-2000 (first entry)
XX
DE Human huntingtin-interacting protein (HIP1).
XX
XX Huntingtin Interacting Protein; HIP; death effector domain; DED; human;
KW apoptosis; HIP-apoptosis modulating protein; cell death; gene therapy;
KW Huntington's disease; neurotropic; anticonvulsant; cyostatic;
KW chromosome 7q11.23.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO960986-A2.
XX
XX PD 02-DEC-1999.
XX
XX PF 27-MAY-1999; 99WO-US11743.
```

```
XX
XX 27-MAY-1998; 98US-0085199.
XX
XX (UYER-) UNIV BRITISH COLUMBIA.
PA (MERI) MERCK FROST CANADA INC.
XX
XX Kalchman M, Hayden MR, Hackam A, Chopra V, Nicholson DM;
PI Vallincoeur JP, Rasper DM;
XX
XX WPI: 2000-097055/08.
DR N-PSDB; AAZ58746.
XX
XX Novel proteins useful for treating Huntington's disease by gene therapy
PT techniques, and cancers -
PT
XX
PS Claim 9; Page 53-57; 91pp: English.
XX
XX The invention relates to Huntingtin Interacting Protein (HIP), that
CC includes a death effector domain (DED), suggesting apoptotic function.
CC Proteins with DED (referred as HIP-apoptosis modulating proteins) are
CC useful for inducing apoptotic death in cells. The HIP is a logical target
CC for therapy in Huntington's disease since it has a differential apoptotic
CC activity, modulated by interaction with Huntingtin having normal and
CC expanded repeats. HIP is also used as a therapeutic agent to introduce
CC apoptosis in cancer cells. Increasing expression of normal (non-expanded)
CC Huntingtin or the HIP-apoptotic modulator-binding portion, a modified
CC HIP-apoptotic modulator in which the DED has been deleted, is helpful for
CC treating Huntington's disease. The present sequence represents the human
CC HIP1.
XX
SQ Sequence 914 AA:

alignment_scores:
Quality: 54.00 Length: 38
Ratio: 2.077 Gaps: 1
Percent Similarity: 68.421 Percent Identity: 36.842

alignment_block:
US-09-684-016-48411/rev x AAY59269 ..
Align seg 1/1 to: AAY59269 from: 1 to: 914

119 GTGTGAATAAATGAGATAGATACCTTACATGGAACCTTACCAG 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 ValPheAsnSerLeuAspMetSerArgSerValSerValThrAlaIaIaG1 95
69 TCAGTGCAGCCCTGCATTCGCCCATTAAGTCATCAGGGTTGTAAGCAT 20
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 yGlnCys.....ArgLeuAlaProLeuIleGlnValIleLeuAspCys 110
19 CTCACATTTTACAT 6
|||||:|||||:|||||:|||||:|||||:|||||:
110 erHisLeuTyTAsp 114

seq_name: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT.AAY59270
seq_documentation_block:
ID AAY59270 standard; Protein; 1090 AA.
XX
XX AAY59270;
XX
DT 17-APR-2000 (first entry)
XX
DE Human huntingtin-interacting protein (HIP1).
XX
XX Huntingtin Interacting Protein; HIP; death effector domain; DED; human;
KW apoptosis; HIP-apoptosis modulating protein; cell death; gene therapy;
KW Huntington's disease; neurotropic; anticonvulsant; cyostatic;
KW chromosome 7q11.23.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO960986-A2.
XX
XX PD 02-DEC-1999.
XX
XX PF 27-MAY-1999; 99WO-US11743.
```







CC secreted protein of the invention.  
 XX  
 SQ Sequence 132 AA;

alignment\_scores:  
 Quality: 53.50 Length: 51  
 Ratio: 2.058 Gaps: 4  
 Percent Similarity: 50.980 Percent Identity: 35.294

alignment\_block:

US-09-684-016-48411/rev x AAE03341 ..

Align seg 1/1 to: AAE03341 from: 1 to: 132

```

116 TGGATAAATTGACAGATGATTCACCTTACATGTGAAA...CTTACCAG 70
||||| ||| |||: |||: ||| ||
50 TrpAnllLeuLeuTyrSerValGlySerIysValSerGlyLeuCysSe 66
69 TCAGTCGACG.....CCTCGCA 53
||:|||||
66 rAsncysSerLeuValProGlyValAlaIAlaHisThrCysAsnProIysV 83
52 TTGGCCCATTAAGTCATCAGGTTGTAG.....GCATCTCAC 15
:: ||||||| |||||||: |||: |||
83 aL...ProLeuGlyLeuGlnGlyCysGluLeuProCysProAlaGluHis 98
14 ATT 12
::
99 Leu 99

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqp/AA12001.DAT:AAE03308

seq\_documentation\_block:

ID AAE03308 standard; Protein: 143 AA.

AC AAE03308;

DT 10-AUG-2001 (first entry)

DE Human gene 17 encoded secreted protein HTOI228, SEQ ID NO:82.

XX Human: secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiotensin disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnerability;  
 KW cell culture; chemotaxis; food additive; gene therapy;  
 KW binding partner identification.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..31 /label= Signal\_peptide

FT Protein 32..132 /label= Human\_mature\_secretcd\_protein

XX WC200134800-A1.

XX 17-MAY-2001.

XX 08-NOV-2000; 2000WO-US30674.

XX 12-NOV-1999; 99US-0164750.

XX 30-JUN-2000; 2000US-0215128.

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX

XX

XX

XX

PI Ruben SM, Komatsoulis GA, Ebner R, Fiscella M, Wei P;  
 XX  
 DR WPI: 2001-329085/34.  
 DR N-PSDB; AAD07721.

PT New nucleic acid molecules encoding human secreted proteins, used in  
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
 PT Parkinson's diseases and cancers -  
 XX  
 PS Claim 11: Page 467-468: 530pp: English.

CC AAD07705-AAD07759 represent cDNAs corresponding to 19 human secreted  
 CC protein genes, and AAE03292-AAE03346 represent the proteins they encode.  
 CC AAE03347-AAE03375 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 19 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis or treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC allergies, neurological disorders (e.g., Alzheimer's disease,  
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
 CC cardiovascular disorders, angiotensin disorders, kidney disorders,  
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
 CC disorders, and infections. The proteins can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues, to identify their  
 CC cognate ligands or binding partners, and in chemotaxis, and can be used  
 CC as a food additive or preservative to modify storage properties.  
 CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 CC secreted protein of the invention.

SQ Sequence 143 AA;

alignment\_scores:  
 Quality: 53.50 Length: 51  
 Ratio: 2.058 Gaps: 4  
 Percent Similarity: 50.980 Percent Identity: 35.294

alignment\_block:

US-09-684-016-48411/rev x AAE03308 ..

Align seg 1/1 to: AAE03308 from: 1 to: 143

```

116 TGGATAAATTGACAGATGATTCACCTTACATGTGAAA...CTTACCAG 70
||||| ||| |||: |||: ||| ||
50 TrpAnllLeuLeuTyrSerValGlySerIysValSerGlyLeuCysSe 66
69 TCAGTCGACG.....CCTCGCA 53
||:|||||
66 rAsncysSerLeuValProGlyValAlaIAlaHisThrCysAsnProIysV 83
52 TTGGCCCATTAAGTCATCAGGTTGTAG.....GCATCTCAC 15
:: ||||||| |||||||: |||: |||
83 aL...ProLeuGlyLeuGlnGlyCysGluLeuProCysProAlaGluHis 98
14 ATT 12
::
99 Leu 99

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV36128

seq\_documentation\_block:



```

ID      AAY36128 standard; Protein; 180 AA.
XX
AC      AAY36128;
XX
DT      13-SEP-1999 (first entry)
XX
DE      Extended human secreted protein sequence, SEQ ID NO. 513.
XX
KW      Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW      cellular differentiation; immune system regulator; anti-inflammatory;
KW      haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW      reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW      genetic disease.
XX
OS      Homo sapiens.
XX
PN      WO9331236-A2.
XX
PD      24-JUN-1999.
XX
PF      17-DEC-1998; 98WO-IB02122.
XX
PR      10-AUG-1998; 98US-0096116.
PR      17-DEC-1997; 97US-0069957.
PR      09-FEB-1998; 98US-0074121.
PR      13-APR-1998; 98US-0081563.
XX
PA      (GEST ) GENSET.
PI      Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR      WPt; 1999-385906/32.
DR      N-PSDB; AAX97812.
XX
PT      New isolated human secreted proteins
XX
PS      Claim 9; Page 447; 516pp; English.
XX
CC      This sequence is encoded by an extended human secreted protein coding
CC      sequence of the invention. The secreted proteins can be used in treating
CC      or controlling a variety of human conditions. The secreted proteins may
CC      act as cytokines or may affect cellular proliferation or differentiation
CC      or may act as immune system regulators, haematopoiesis regulators, tissue
CC      growth regulators, regulators of reproductive hormones or cell movement
CC      or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC      tumour inhibition activity. The DNAs can be used in forensic procedures
CC      to identify individuals or in diagnostic procedures to identify
CC      individuals having genetic diseases resulting from abnormal expression of
CC      the genes corresponding to the extended cDNAs. They are also useful for
CC      constructing a high resolution map of the human chromosomes. They can
CC      also be used for gene therapy to control or treat genetic diseases.
XX
Sequence   180 AA;

Alignment_scores:
          Quality:    53.00      Length:    28
          Ratio:     2.650       Gaps:      0
Percent Similarity:  71.429     Percent Identity: 42.857

Alignment_block:
US-09-684-016-48411 x AAY36128 ..

Align seg 1/1 to: AAY36128 from: 1 to: 180

38 GACCTAATGGCGCAATGCCAGGCGTCGACACGACGTGAAGTTTCACATCT 87
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69 ASPLIErThALaspleugiuaSnIlIephEasPrIpasnaVallySgInLe 85
88 AAGGTGATACTATTCTTCATTTTATTCACACACA 121
: ||||||||| : ||||||||| :
85 upHeUrYrleuSerAlaGlutyrSerThrLys 96

```

```

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT: AAB24029
seq_documentation_block:
ID AAB24029 standard; Protein: 180 AA.
xx
xx
AC AAB24029;
xx
DT 25-JAN-2001 (first entry)
xx
DE Human PRO3567 protein sequence SEQ ID NO:27.
xx
KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KW identification; tumourigenesis; anticancer; detection.
xx
OS Homo sapiens.
xx
PN WO200053750-A1.
xx
PD 14-SEP-2000.
xx
PF 02-DEC-1999; 99WO-US28551.
xx
PR 08-MAR-1999; 99WO-US05028.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
xx
PA (GETH ) GENENTECH INC.
PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
xx
DR WPI: 2000-594320/56.
DR N-PSDB; AAC58111.
xx
PT Antibodies specific for PRO polypeptides, used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of PRO
PT polypeptide activity or expression -
xx
PS Claim 61; Fig 18; 226pp; English.
xx
xx
The present invention describes an antibody that binds to a human
protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
PRO3434; PRO1927; PRO3567; PRO1295; PRO1293; PRO1303; PRO3444; PRO3544;
PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
anticancer activity and can be used to diagnose tumors in mammals, by
detecting complex formation when the antibody is contacted with test
cells. Increased expression of genes encoding (I) can also be detected
to diagnose tumors. Agents which inhibit the activity of (I),
especially the antibodies, or an antisense oligonucleotide which
hybridises to genes encoding (I), can be used to inhibit tumour growth,
preferably by inducing cell death. Methods from the present invention
can be used to identify compounds which inhibit the biological activity
of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
probes used in examples from the present invention for human PRO
sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
PRO polynucleotide and protein sequences given in the exemplification of
the present invention.
xx
SO Sequence 180 AA;

alignment_scores:
Quality: 53.00 Length: 28
Ratio: 2.650 Gaps: 0
Percent Similarity: 71.429 Percent Identity: 42.857

alignment_block:
US-09-684-016-48411 x AAB24029 ..
Align seg 1/1 to: AAB24029 from: 1 to: 180

```







CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
 CC are used (i) to identify agents suitable for treatment of uterine or  
 CC endometrial cancer; (ii) directly for treating these forms of cancer  
 CC (including expression from gene therapy vectors) and (iii) for  
 CC generation of specific antibodies. (A) are identified by assembling ESTs  
 CC (expressed sequence tags) from a particular tissue type before comparison  
 CC of expression patterns. This allows a significantly longer fragment of  
 CC the gene to be revealed, so should reduce the number of failures  
 CC associated with the fact that ESTs from different libraries may represent  
 CC different parts of the same unknown gene, distorting the estimated  
 CC frequency of occurrence in a particular tissue. AAY5941-Y60328 represent  
 CC protein fragments encoded by the human endometrium tumour cDNA library  
 CC derived EST fragments represented in AAZ41981-742121.

XX Sequence 66 AA;

alignment\_scores:                   Quality: 51.00                   Length: 30  
                                   Ratio: 2.550                   Gaps: 1  
                                   Percent Similarity: 66.667           Percent Identity: 33.333

alignment\_block:  
 US-09-684-016-48411/rev x AAY60285

Align seg 1/1 to: AAY60285 from: 1 to: 66

```

110 AATTGAGATAGTATCATCTTACATGTG.....AACTTACCACTCA 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 LysIleSerLeuSerPheSerIleuHisIleMetPheIleGlnAArgH1 48
66 GTGCAGCCCTCCGATTCGCCCATTTAGCTCATTCAGGCTTGT 27
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
48 sCysArgGluArgValArgProCysGlyGluLeuMetCys 61

```

seq\_name: /SID52/gcgcdata/geneseq/geneseqp/AA1997.DAT:AAW20131

seq\_documentation\_block:

ID AAW20131 standard; Protein: 288 AA.

XX AAW20131;

DT 08-JUL-1997 (first entry)

DE H. pylori cell envelope transporter protein 13726562.aa.

XX Cytoplasmic; vaccine: prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT Misc-difference 8

FT /label= unknown  
 FT /note= "encoded by YAA"

XX W09640893-A1.

XX 19-DEC-1996.

XX PF 06-JUN-1996; 96WO-US09122.

XX PR 01-APR-1996; 96US-0630405.

XX PR 07-JUN-1995; 95US-0487032.

XX PA (ASTR ) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaerd BL;

XX WPI: 1997-052306/05.

XX DR N-PSDB; AAT67376.

PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter  
 XX Claim 56: Page 356; 1481pp; English.  
 XX The present sequence shows a Helicobacter pylori cell envelope  
 CC transporter protein.  
 CC This protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.

XX Sequence 288 AA;

alignment\_scores:                   Quality: 51.00                   Length: 31  
                                   Ratio: 2.684                   Gaps: 0  
                                   Percent Similarity: 61.290           Percent Identity: 38.710

alignment\_block:  
 US-09-684-016-48411/rev x AAW20131

Align seg 1/1 to: AAW20131 from: 1 to: 288

```

110 AATTGAGATAGTATCATCTTACATGTGAACTTACCACTGACGAGCGAG 61
   ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
156 LysSerIleHisProValLeuLysAsnIleAsnLeuThrIleGlnLysC1 172
60 CCCTGCATTCGCCCATTTAGCTCATTCAGGCTTGTAGGCATCT 18
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:
172 YGlnIleLeuLeuLeuIleGlyHisSerGlyCysGlyLysSer 186

```

seq\_name: /SID52/gcgcdata/geneseq/geneseqp/AA1997.DAT:AAW20619

seq\_documentation\_block:

ID AAW20619 standard; Protein: 377 AA.

XX AAW20619;

DT 14-JUL-1997 (first entry)

DE H. pylori transporter protein, 02ae1161lori11.

XX Cytoplasmic; vaccine: prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis.

XX Helicobacter pylori.

XX W09640893-A1.

XX 19-DEC-1996.

XX PF 06-JUN-1996; 96WO-US09122.

XX PR 01-APR-1996; 96US-0630405.

XX PR 07-JUN-1995; 95US-0487032.

XX PA (ASTR ) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaerd BL;



XX WPI: 1997-052306/05.  
DR N-PSDB: AAT67872.

XX Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter

XX PS Claim 56: Page 1045-46; 1481pp; English.

CC This sequence represents a H. pylori transporter protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.

XX Sequence 377 AA;

alignment\_scores:  
Quality: 51.00 Length: 31  
Ratio: 2.684 Gaps: 0  
Percent Similarity: 61.290 Percent Identity: 38.710

alignment\_block:

US-09-684-016-48411/rev x AAW20619 ..

Align seg 1/1 to: AAW20619 from: 1 to: 377

```

110 AATGTGAGATGAATGATGACCTTACATGGAAGTACTACAGTCAGTCAG 61
    ||| |||   ::   ::   ::   ||| |||   ::
174 LysSerLysHisProValLeuLysAsnPheAsnLeuThrIleGlnLysG1 190
    : ::||| |||   ::||| |||   ::||| |||   |||
60 CCCTGCGCATGGCCCATTAAGTCATCAGGGGTGTGAAGGCATCT 18
    : ::||| |||   ::||| |||   ::||| |||   |||
190 yGlnLysIleAlaLeuIleGlyHisSerGlyCysGlyLysSer 204

```



**THIS PAGE BLANK (USPTO)**



us-09-684-016-48411.std.raii

/cgn2_6/ptodata/2/1aa/55a.COMB	pep:US-08-484-496-14	-	47.00	101.95	54.28	58.58
/cgn2_6/ptodata/2/1aa/55b.COMB	pep:US-08-345-212-14	-	47.00	101.95	54.28	58.58
/cgn2_6/ptodata/2/1aa/63b.COMB	pep:US-09-449-033-14	-	47.00	101.95	54.28	58.58
/cgn2_6/ptodata/2/1aa/55b.COMB	pep:US-08-374-483-2	+	47.00	98.74	58.28	820
/cgn2_6/ptodata/2/1aa/55b.COMB	pep:US-08-374-483-5	+	47.00	98.74	58.28	820

seq\_name: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:US-08-762-428A-6

```

:seq_name: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:US-08-762-428A-6
:seq_documentation_block:
:  Sequence 6, Application US/08762428A
:  Patent No. 6120993
:  GENERAL INFORMATION:
:  APPLICANT: Ye, Guo-Jie
:  APPLICANT: Breslow, Esther M.
:  APPLICANT: Meister, Alton
:  TITLE OF INVENTION: 5-OXOPROLINASE
:  NUMBER OF SEQUENCES: 10
:  CORRESPONDENCE ADDRESS:
:  ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
:  STREET: Clinton Square, P.O. Box 1051

```

```

; STATE: New York
; COUNTRY: USA
;

```

```

?
?      ZIP:    14805
?
?      COMPUTER READABLE FORM:
?
?      MEDIUM TYPE: Floppy disk
?
?      COMPUTER: IBM PC compatible
?
?      OPERATING SYSTEM: PC-DOS/MS-DOS
?
```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/762,428A
: FILING DATE:
:

```

```

1 CLASSIFICATION: 536
2 ATTORNEY/AGENT INFORMATION:
3 NAME: TIMIAN, SUSAN J.
4 REGISTRATION NUMBER: 34,103
5 REFERENCE/DOCKET NUMBER: 19603/1330
6 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: 716-263-1636
8 TELEFAX: 716-263-1600
9 INFORMATION FOR SEQ ID NO: 6:
10 SEQUENCE CHARACTERISTICS:
11 LENGTH: 1288 amino acids
12 TYPE: amino acid
13 STRANDEDNESS: not relevant
14 TOPOLOGY: linear
15 MOLECULE TYPE: protein
16 US-08-762-428A-6
17
18 alignment_scores:
19 Quality: 56.00 Length: 30
20 Ratio: 2.800 Gaps: 1
21 Percent Similarity: 66.667 Percent Identity: 40.000
22
23 alignment_block:
24 US-09-684-016-48411/rev x US-08-762-428A-6 ..
25
26 Align seg 1/1 to: US-08-762-428A-6 from: 1 to: 1288
27
28 86 CATGTGAACCTTACCGATCGAGTCAGCCCT.....CGCATTTGGCCATT 43
29 |||||:||||:||||: || |||||:||||
30 |||||:||||:||||: || |||||:||||
31 215 H1SVLserLseuSerSerGluValMetProMetValArgIleValProAr 231
32
33 42 AGGTATCATCGGCTGTGAAGCATCTCACATTTACATATCC 3
34 ||||| :||| ||| :|||
35 ||||| :||| ||| :|||
36 231 gGlyHisThrAlaCysAlaAspAlaTyLeuThrProThr 244
37
38 seq_name: /cgn2_6/plodata/2/iaa/5B_COMB.pep:US-09-085-199B-4
39
40 seq_documentation_block:
41 Sequence 4, Application US/09085199B
42 Patent No. 6235879

```



; GENERAL INFORMATION

GENERAL INFORMATION:



```

; APPLICANT: Loosmore, Sheena M
; APPLICANT: Klein, Michael H
; TITLE OF INVENTION: Haemophilus Influenzae
; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,515
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-549-515-9

alignment_scores:
      Quality: 51.00      Length: 29
      Ratio: 3.400      Gaps: 1
      Percent Similarity: 51.724      Percent Identity: 34.483

alignment_block:
US-09-684-016-48411 x US-08-549-515-9  ..
Align seg 1/1 to: US-08-549-515-9 from: 1 to: 207

      2 AGTATGTGAATGTGAGATGCTTACACCGCTGATGACCTA..... 43
      |||||||      |||||||      ::
      103 ArgTyrCysHisMetAlaCysProTyrGlyAlaProGlnTyrAsnGluTh 119
      44 .....ATGGCGCAATGCCAGGCGCTGCACCTGAC 70
      |||:|||||:|||||:|||||:||||| 111
      119 rLysGlnHisMetThrLysCysAspGlyCysTyrAsp 131

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-866-545-4

seq_documentation_block:
; Sequence 4, Application US/08866545
; Patent No. 6265535
; GENERAL INFORMATION:
; APPLICANT: Greene, Mark I.
; APPLICANT: Murall, Ramachandran
; APPLICANT: Takasaki, Wataru
; TITLE OF INVENTION: PEPTIDES AND PEPTIDE
; TITLE OF INVENTION: ANALOGUES DESIGNED FROM BINDING SITES OF TUMOR
; TITLE OF INVENTION: NECROSIS FACTOR RECEPTOR SUPERFAMILY AND THEIR
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York

```

```

; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,545
; FILING DATE: 30-MAY-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 009113-0004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: NO. 6265535e
;
US-08-866-545-4

alignment_scores:
      Quality: 50.00      Length: 19
      Ratio: 4.167      Gaps: 0
      Percent Similarity: 63.158      Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-866-545-4  ..
Align seg 1/1 to: US-08-866-545-4 from: 1 to: 76

      8 TGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCGCAATGCGA 57
      |||:|||||      |||:|||||      :|||:|||||
      44 CysArgCysAlaTyrGlyTyrTyrGlnAspGluThrThrGlyArgCysG1 60
      58 GGCGCTGC 64
      ||:|||||
      60 uAlaCys 62

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-232-087A-11

seq_documentation_block:
; Sequence 11, Application US/08232087A
; Patent No. 5866372
; GENERAL INFORMATION:
; APPLICANT: Stein, Harald
; APPLICANT: Dikop, Horst
; APPLICANT: Latza, Ute
; TITLE OF INVENTION: Lymphoid CD30-Antigen
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS

```



```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/732,087A
FILING DATE: 08-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 736-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Protein
LOCATION: 1..159
OTHER INFORMATION: /note="NGFR, see Fig. 5"
US-08-232-087A-11

```

```

alignment_scores:
Quality: 50.00 Length: 20
Ratio: 3.846 Gaps: 0
Percent Similarity: 65.000 Percent Identity: 35.000

```

```

alignment_block:
US-09-684-016-48411 x US-08-232-087A-11 ..

```

```

Align seg 1/1 to: US-08-232-087A-11 from: 1 to: 159

```

```

8 TGTAAATGTGAGATGCTTACACCTGATGACCTAATGCGCATGCGA 57
||||:||||| ||| |||:|:|||||
77 CysArgCysAlaTyrGlyTyrTyrGlnAspGluThrThcGlyArgCysG1 93
58 GGGCTGCGACT 67
|:|:|:|:|:|:|
93 uAlaCysSer 96

```

```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-219-237B-6

```

```

seq_documentation_block:
Sequence 6, Application US/08219237B
Patent No. 5874546
GENERAL INFORMATION:
APPLICANT: NAGATA, Shigekazu
APPLICANT: ITOH, Naoto
APPLICANT: YONEHARA, Shin
TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: James W. Hellwege
STREET: P.O. Box 2266 Eads Station
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, G. Kevin
APPLICATION NUMBER: US/08/219,237B
FILING DATE: 28-MAR-1994

```

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,129
FILING DATE: 22-APR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: James W. Hellwege
REGISTRATION NUMBER: 28,808
REFERENCE/DOCKET NUMBER: 516762
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-219-237B-6

```

```

alignment_scores:
Quality: 50.00 Length: 19
Ratio: 4.167 Gaps: 0
Percent Similarity: 63.158 Percent Identity: 36.842

```

```

alignment_block:
US-09-684-016-48411 x US-08-219-237B-6 ..

```

```

Align seg 1/1 to: US-08-219-237B-6 from: 1 to: 159

```

```

8 TGTAAATGTGAGATGCTTACACCTGATGACCTAATGCGCATGCGA 57
||||:||||| ||| |||:|:|||||
77 CysArgCysAlaTyrGlyTyrTyrGlnAspGluThrThcGlyArgCysG1 93
58 GGGCTGCG 64
|:|:|:|:|
93 uAlaCys 95

```

```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-477-347-15

```

```

seq_documentation_block:
Sequence 15, Application US/08477347
Patent No. 6232446
GENERAL INFORMATION:
APPLICANT: WALLACH, David
APPLICANT: BIGDA, Jacek
APPLICANT: BELETSKY, Igor
APPLICANT: METT, Igor
TITLE OF INVENTION: TNF LIGANDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,347
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/115,685
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 106271
FILING DATE: 08-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, G. Kevin
APPLICATION NUMBER: 34,033
REGISTRATION NUMBER:

```



```

: REFERENCE/DOCKET NUMBER: WALLACH=10
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 159 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-477-347-15

alignment_scores:
      Quality: 50.00      Length: 19
      Ratio: 4.167      Gaps: 0
      Percent Similarity: 63.158      Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-477-347-15 ..

Align seg 1/1 to: US-08-477-347-15 from: 1 to: 159

      8 TTTAATGTGAGATGCGCTTACAAACCTGATGACCTTAATGGCCCAATGCGA 57
      |||:::||||| ||| |||::: |||::: |||::: |||::: |||::: |||
      77 GYSAAGCYSAIATyTgIyTfTyTgInaspGIuThrThgIyArGcysGsl 93
      93 uAlAcys 95

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-476-862-6

seq_documentation_block:
: Sequence 6, Application US/08476862
: Patent No. 6262239
: GENERAL INFORMATION:
: APPLICANT: WALLACH, David
: APPLICANT: BIGDA, Jacek
: APPLICANT: BELITSKY, Igor
: APPLICANT: METT, Igor
: APPLICANT: ENGELMANN, Hartmut
: TITLE OF INVENTION: TNF INHIBITORS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROMY AND NEIMARK
: STREET: 419 Seventh Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/476,862
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: IL 107267
: FILING DATE: 12-OCT-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: IL 94039
: FILING DATE: 06-APR-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: IL 91229
: FILING DATE: 06-AUG-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: IL 90339

```

```

      FILING DATE: 18-MAY-1989
      ATTORNEY/AGENT INFORMATION:
        NAME: BROWDY, Roger L.
        REGISTRATION NUMBER: 25,618
        REFERENCE/DOCKET NUMBER: WALLACH-12A
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: 202-628-5197
          TELEFAX: 202-737-3528
        INFORMATION FOR SEQ ID NO: 6:
          SEQUENCE CHARACTERISTICS:
            LENGTH: 159 amino acids
            TYPE: amino acid
            STRANDEDNESS: single
            TOPOLOGY: linear
          MOLECULE TYPE: peptide
          US-08-476-862-6

alignment_scores:
      Quality: 50.00      Length: 19
      Ratio: 4.167      Gaps: 0
      Percent Similarity: 63.158      Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-476-862-6 ..

Align seg 1/1 to: US-08-476-862-6 from: 1 to: 159

      8 TCTAATGTGAGATGGCTTACAAACCTGATGACCTAATGACCGCAATGCGA 57
      |||:||| ||| |||:|:|:|
      77 CysATGcGsaIatYrGcIyTyrTyrGlnaspGluThrThrGlyArgCysel 93
      58 GGGCTGC 64
      |::|||
      93 uLaCys 95

seq_name: /cgn2_6/ptcdatdata/2/1aa/6B_COMB.dep:US-08-468-560C-6

seq_documentation_block:
  Sequence 6, Application US/08468560C
  Patent No. 6270998
  GENERAL INFORMATION:
    APPLICANT: NAGATA, Shigekazu
    APPLICANT: ITOH, Naoto
    APPLICANT: YONEMURA, Shin
    TITLE OF INVENTION: DNA CODING FOR HUMAN CELL SURFACE
    TITLE OF INVENTION: ANTIGEN
    NUMBER OF SEQUENCES: 11
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP.
      STREET: P.O. BOX 747
      CITY: FALLS CHURCH
      STATE: VA
      COUNTRY: USA
      ZIP: 22040-0747
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/468,560C
      FILING DATE: 06-JUN-1995
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
        NAME: MURPHY JR., GERRAD M.
        REGISTRATION NUMBER: 28,977
        REFERENCE/DOCKET NUMBER: 20-4393P
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: 703-205-8000
          TELEFAX: 703-205-8050
      INFORMATION FOR SEQ ID NO: 6:
        SEQUENCE CHARACTERISTICS:

```



LENGTH: 159 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-468-560C-6

alignment\_scores:  
Quality: 50.00 Length: 19  
Ratio: 4.167 Gaps: 0  
Percent Similarity: 63.158 Percent Identity: 36.842

alignment\_block:

US-09-684-016-48411 x US-08-468-560C-6 ..

Align seg 1/1 to: US-08-468-560C-6 from: 1 to: 159

8 TGTAAATGTGAGATGCCTTACACCTGATGACCTAATGGCGCAATGCCA 57  
||||:|||||  
77 CysArgCysAlaTyrGlyTyrGlnAspGluThrThrGlyArgCysG1 93  
58 GGGCTGC 64  
|:::|  
93 uAlaCys 95

seq\_name: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:US-08-974-022-50

seq\_documentation\_block:

Sequence 50, Application US/08974022  
Patent No. 6015938  
GENERAL INFORMATION:  
APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974.022  
FILING DATE: 12-DEC-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-974-022-50

alignment\_scores:  
Quality: 50.00  
Ratio: 4.167

Length: 19  
Gaps: 0

Percent Similarity: 63.158 Percent Identity: 36.842

alignment\_block:

US-09-684-016-48411 x US-08-974-022-50 ..

Align seg 1/1 to: US-08-974-022-50 from: 1 to: 224

8 TGTAAATGTGAGATGCCTTACACCTGATGACCTAATGGCGCAATGCCA 57  
||||:|||||  
107 CysArgCysAlaTyrGlyTyrGlnAspGluThrThrGlyArgCysG1 123  
58 GGGCTGC 64  
|:::|  
123 uAlaCys 125

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-08-795-445A-50

seq\_documentation\_block:

Sequence 50, Application US/08795445A  
Patent No. 6284485  
GENERAL INFORMATION:  
APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795.445A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-445A-50

alignment\_scores:  
Quality: 50.00 Length: 19  
Ratio: 4.167 Gaps: 0  
Percent Similarity: 63.158 Percent Identity: 36.842

alignment\_block:

US-09-684-016-48411 x US-08-795-445A-50 ..

Align seg 1/1 to: US-08-795-445A-50 from: 1 to: 224

8 TGTAAATGTGAGATGCCTTACACCTGATGACCTAATGGCGCAATGCCA 57  
||||:|||||  
107 CysArgCysAlaTyrGlyTyrGlnAspGluThrThrGlyArgCysG1 123



```
58 GGCCTGC 64
|:::|
123 uAlacys 125

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-795-447A-50
seq_documentation_block:
; Sequence 50, Application US/08795447A
; Patent No. 6284728
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: Osteoprotegerin
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: One Amgen Center Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91362-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,447A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378D2
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-795-447A-50

alignment_scores:
Quality: 50.00 Length: 19
Ratio: 4.167 Gaps: 0
Percent Similarity: 63.158 Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-795-447A-50 ..

Align seq 1/1 to: US-08-795-447A-50 from: 1 to: 224

8 TGTAAATGTGATGCTTACACCGCTGATGACCTAATGGCGCATGCGA 57
|:::| |:::| |:::|
107 CysarGcysAlaTyrglYtYrGlmspGluhrthrglYArGcysgl 123
58 GGCCTGC 64
|:::|
123 uAlacys 125

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
; Sequence 50, Application US/08974186
; Patent No. 6284740
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi

; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

alignment_scores:
Quality: 50.00 Length: 19
Ratio: 4.167 Gaps: 0
Percent Similarity: 63.158 Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-974-186-50 ..

Align seq 1/1 to: US-08-974-186-50 from: 1 to: 224

8 TGTAAATGTGATGCTTACACCGCTGATGACCTAATGGCGCATGCGA 57
|:::| |:::| |:::|
107 CysarGcysAlaTyrglYtYrGlmspGluhrthrglYArGcysgl 123
58 GGCCTGC 64
|:::|
123 uAlacys 125

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-795-446B-50
seq_documentation_block:
; Sequence 50, Application US/08795446B
; Patent No. 6288032
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,446B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-446B-50

```

```

alignment_scores:
Quality: 50.00 Length: 19
Ratio: 4.167 Gaps: 0
Percent Similarity: 63.158 Percent Identity: 36.842

```

## alignment\_block:

```
US-09-684-016-48411 x US-08-795-446B-50 ..
```

```
Align seg 1/1 to: US-08-795-446B-50 from: 1 to: 224
```

```

8 TGTAAATGTGATGCTTACAAACCTGATGACCTATGCGCGCATGCGA 57
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 CysArGcysAlaTyrGlyTyrGlnAspGluThrGlyArgCysgl 123

```

```
58 GGGCTGC 64
```

```
123 uAlAcYs 125
```

```
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-864-804-1
```

```

seq_documentation_block:
; Sequence 1, Application US/08864804
; Patent No. 5948641
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Puri
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: METAL RESPONSE ELEMENT BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,804
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.

```

```

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0307 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LIVERPOT2
CLONE: 2048959
US-08-864-804-1

```

```

alignment_scores:
Quality: 50.00 Length: 23
Ratio: 3.571 Gaps: 0
Percent Similarity: 60.870 Percent Identity: 39.130

```

## alignment\_block:

```
US-09-684-016-48411 x US-08-864-804-1 ..
```

```
Align seg 1/1 to: US-08-864-804-1 from: 1 to: 420
```

```

5 TATTGTAAATGTGATGCTTACAAACCTGATGACCTATGCGCGCATG 54
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 TyrcysTyrCysGlyGlyProGlyGluTrpAsnLeuLysMetLeuGlnCys 205

```

```
55 CGAGGGCTGCACCTGACTG 73
```

```
205 sarGserCysLeuGlnTrp 211
```

```
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-086-483A-4
```

```

seq_documentation_block:
; Sequence 4, Application US/09086483A
; Patent No. 6214580
; GENERAL INFORMATION:
; APPLICANT: NI, et al.
; TITLE OF INVENTION: HUMAN TUMOR NECROSIS FACTOR RECEPTOR TRIO
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/086,483A
; FILING DATE: May-29-98
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/050,936
; FILING DATE: May-30-97
; CLASSIFICATION: 435
; APPLICATION NUMBER: 60/069,112
; FILING DATE: Dec-9-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A.
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF379

```



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 427 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-086-483A-4

alignment\_scores:  
Quality: 50.00 Length: 19  
Ratio: 4.167 Gaps: 0  
Percent Similarity: 63.158 Percent Identity: 36.842

## alignment\_block:

US-09-684-016-48411 x US-09-086-483A-4 ..

Align seg 1/1 to: US-09-086-483A-4 from: 1 to: 427

8 TGTAAATGTGAGATGCTTACACCGTGATGACCTTAATGGCGCAATGCGA 57  
||||:||||| ||| |||:|  
107 CysArgCysAlaTyrGlyTyrTyrGlnAspLurThrThrGlyArgCysG1 123  
58 GGCGCTGC 64  
|:|:|  
123 uAlaCys 125

seq\_name: /cgn2\_6/ptodata/2/laa/6B\_COMB.pep:US-09-041-886-2

## seq\_documentation\_block:

Sequence 2, Application US/09041886  
Patent No. 6235872  
GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 427 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-041-886-2

alignment\_scores:  
Quality: 50.00 Length: 19  
Ratio: 4.167 Gaps: 0  
Percent Similarity: 63.158 Percent Identity: 36.842

## alignment\_block:

US-09-684-016-48411 x US-09-041-886-2 ..

Align seg 1/1 to: US-09-041-886-2 from: 1 to: 427

8 TGTAAATGTGAGATGCTTACACCGTGATGACCTTAATGGCGCAATGCGA 57  
||||:||||| ||| |||:|  
107 CysArgCysAlaTyrGlyTyrTyrGlnAspLurThrThrGlyArgCysG1 123  
58 GGCGCTGC 64  
|:|:|  
123 uAlaCys 125

seq\_name: /cgn2\_6/ptodata/2/laa/6B\_COMB.pep:US-09-006-353A-5

## seq\_documentation\_block:

Sequence 5, Application US/09006353A  
Patent No. 6261801  
GENERAL INFORMATION:  
APPLICANT: WEI, YING-FEI  
APPLICANT: GENTZ, REINER  
APPLICANT: RUBEN, STEVEN  
TITLE OF INVENTION: TUMOR NECROSIS FACTOR RECEPTOR 5  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: US  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/006,353A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOKES, ANDERS A.  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF341  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8512  
TELEFAX: (301) 309-8504  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 427 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-006-353A-5

alignment\_scores:  
Quality: 50.00 Length: 19  
Ratio: 4.167 Gaps: 0  
Percent Similarity: 63.158 Percent Identity: 36.842

## alignment\_block:

US-09-684-016-48411 x US-09-006-353A-5 ..

Align seg 1/1 to: US-09-006-353A-5 from: 1 to: 427



8 TGTAAATGAGATGCTTACCAACCTGATGACCTAATGGCGCATGCGA 57  
 |||:::|||||  
 107 CysATGCGyAlATyTgTyTgTgInAspIuThrTgIyAtrGysG1 123  
 58 GGGCTGC 64  
 |||:::|||||  
 123 uAlAcys 125

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-553-436A-8

seq\_documentation\_block:

; Sequence 8, Application US/08553436A  
 ; Patent No. 5866790  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HESSE, Holger  
 ; APPLICANT: MULLER-ROBER, Bernd  
 ; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE  
 ; TITLE OF INVENTION: PREPARATION OF SUGAR BEET WITH CHANGED SUCROSE  
 ; TITLE OF INVENTION: CONCENTRATION  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
 ; STREET: 1180 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: US  
 ; ZIP: 10036-8403  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/553.436A  
 ; FILING DATE: 17-NOV-1995  
 ; CLASSIFICATION: 800  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/EP94/01671  
 ; FILING DATE: 20-MAY-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: DE P 4317596.1  
 ; FILING DATE: 24-MAY-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meilman, Edward  
 ; REGISTRATION NUMBER: 24,735  
 ; REFERENCE/DOCKET NUMBER: P/951-117  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 382-0700  
 ; TELEFAX: (212) 382-0888  
 ; TELEX: 236925  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 766 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-553-436A-8

alignment\_scores:

Quality:	50.00	Length:	57
Ratio:	2.273	Gaps:	2
Percent Similarity:	38.596	Percent Identity:	26.316

alignment\_block:

US-09-684-016-48411/rev x US-08-553-436A-8 ..

Align seq 1/1 to: US-08-553-436A-8 from: 1 to: 766

105 GAAGATAGTATACCTTACATGTGAA..... 79  
 |||||||||||:|||||:  
 399 GluAspLysTyrHisPheSerCysGlnPheSerAlaAspLeuMetAlaLeu 415

78 .....ACTTACCACTGAGT..... 64  
 |||:::|||||  
 415 tAsnH1AlAspPheIleIleThrSerThrTgInGluIleAlaIGLYT 432  
 63 .....CAGCCCTGCGATTCGCCATTAAGTCATCA 34  
 |||:::|||||  
 432 hrlYsAsnThrValGlyGlnTyrGluSerHisLysAlaPheThrPhePro 448  
 33 GGGTGTAGGACATTCACAT 13  
 |||:::|||||  
 449 GlyLeuTyrArgValValHis 455

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-453-472-6

seq\_documentation\_block:

; Sequence 6, Application US/08453472  
 ; Patent No. 5626846  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DEAN, JURRIEN  
 ; TITLE OF INVENTION: CONTRACEPTIVE VACCINE  
 ; TITLE OF INVENTION: BASED ON ALLOIMMUNIZATION WITH ZONA PELLUCIDA  
 ; TITLE OF INVENTION: POLYPEPTIDES  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: MORGAN & FINNEGAN  
 ; STREET: 345 PARK AVENUE  
 ; CITY: NEW YORK  
 ; STATE: NEW YORK  
 ; COUNTRY: USA  
 ; ZIP: 10154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: FLOPPY DISK  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/453.472  
 ; FILING DATE: 30-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/038,948  
 ; FILING DATE: 26-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/930,462  
 ; FILING DATE: 20-AUG-1992  
 ; APPLICATION NUMBER: US 07/364,379  
 ; FILING DATE: 12-JUN-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: DOROTHY R. AOTH  
 ; REGISTRATION NUMBER: 36,434  
 ; REFERENCE/DOCKET NUMBER: 2026-4032 US3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 758-4800  
 ; TELEFAX: (212) 751-6849  
 ; TELEX: 421792  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 713  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: mouse  
 ; STRAIN:  
 ; INDIVIDUAL ISOLATE:  
 ; DEVELOPMENTAL STAGE:  
 ; HAPLOTYPE:  
 ; TISSUE TYPE:  
 ; CELL TYPE:  
 ; CELL LINE:











```
SEQUENCE CHARACTERISTICS:
LENGTH: 713
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: mouse
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL LINE:
ORGANELLE:
FEATURE:
NAME/KEY: ZP2
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: mouse ZP2 protein
US-08-453-952-6

alignment_scores:
Quality: 48.50 Length: 38
Ratio: 2.553 Gaps: 2
Percent Similarity: 50.000 Percent Identity: 34.211

alignment_block:
US-09-684-016-48411 x US-08-453-952-6 ..
Align seg 1/1 to: US-08-453-952-6 from: 1 to: 713

8 TGTAAATGTGAGATGCTTACACACCTGATGACCTAATGGCGCAATGCGA 57
||| ||||| ||| :: ||::||| ||||| ::
353 CysHiscysgluserProvalSerIleaspIleucysalagin...As 368
58 GGGCTGCACGTGCTGTAAGTTTCACATGTAAGTGATCTACTTCA 107
||| ||| ::|
368 pGlyPheMetasphe.....GluV 375

108 TTATTCACACAA 121
::||| |||
375 alTyrSerHisGln 379

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-862-903-6

seq_documentation_block:
; Sequence 6, Application US/08862903
; Patent No. 5916768
; GENERAL INFORMATION:
; APPLICANT: DEAN, JURRIEN
; TITLE OF INVENTION: CONTRACEPTIVE VACCINE
; TITLE OF INVENTION: BASED ON ALLOIMUNIZATION WITH ZONA PELLUCIDA
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,903
; FILING DATE: 30-May-1995
; CLASSIFICATION: 424
```

```
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/038,948
FILING DATE: 26-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,462
FILING DATE: 20-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/364,379
FILING DATE: 12-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36,434
REFERENCE/DOCKET NUMBER: 2026-4032 USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 713
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: mouse
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL LINE:
ORGANELLE:
FEATURE:
NAME/KEY: ZP2
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: mouse ZP2 protein
US-08-862-903-6

alignment_scores:
Quality: 48.50 Length: 38
Ratio: 2.553 Gaps: 2
Percent Similarity: 50.000 Percent Identity: 34.211

alignment_block:
US-09-684-016-48411 x US-08-862-903-6 ..
Align seg 1/1 to: US-08-862-903-6 from: 1 to: 713

8 TGTAAATGTGAGATGCTTACACACCTGATGACCTAATGGCGCAATGCGA 57
||| ||||| ||| :: ||::||| ||||| ::
353 CysHiscysgluserProvalSerIleaspIleucysalagin...As 368
58 GGGCTGCACGTGCTGTAAGTTTCACATGTAAGTGATCTACTTCA 107
||| ||| ::|
368 pGlyPheMetasphe.....GluV 375

108 TTATTCACACAA 121
::||| |||
375 alTyrSerHisGln 379

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-549-515-6

seq_documentation_block:
; Sequence 6, Application US/08549515
; Patent No. 6054123
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Haemophilus Influenzae
```



```

; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,515
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-549-515-6

alignment_scores:
      Quality: 48.00      Length: 27
      Ratio: 3.429      Gaps: 1
      Percent Similarity: 51.852      Percent Identity: 33.333

alignment_block:
US-09-684-016-48411 x US-08-549-515-6  ..

Align seg 1/1 to: US-08-549-515-6 from: 1 to: 205

2 AGGATGTGTAATGTGAGATGCTTACAACCTGATGACCTA..... 43
||||||||| |||||||:::
103 ArgTYrCYshISmetAlaCYsProTYAspAlaPrGlnTYrAspAlaG1 119
44 .....ATGGCGCAATGCGAGGCGCTGC 64
|||||:::|||||:::|||||
119 nLYsgLYhISmetThrLYsCYsAspGLYcys 129

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-549-515-12

seq_documentation_block:
; Sequence 12, Application US/08549515
; Patent No. 6034123
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Kleiu, Michel H
; TITLE OF INVENTION: Haemophilus Influenzae
; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,515
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-549-515-12

alignment_scores:
      Quality: 48.00      Length: 27
      Ratio: 3.429      Gaps: 1
      Percent Similarity: 51.852      Percent Identity: 33.333

alignment_block:
US-09-684-016-48411 x US-08-549-515-12  ..

Align seg 1/1 to: US-08-549-515-12 from: 1 to: 205

2 AGGATGTGTAATGTGAGATGCTTACAACCTGATGACCTA..... 43
||||||||| |||||||:::
103 ArgTYrCYshISmetAlaCYsProTYAspAlaPrGlnTYrAspAlaG1 119
44 .....ATGGCGCAATGCGAGGCGCTGC 64
|||||:::|||||:::|||||
119 nLYsgLYhISmetThrLYsCYsAspGLYcys 129

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-249-112-4

seq_documentation_block:
; Sequence 4, Application US/08249112
; Patent No. 5527703
; GENERAL INFORMATION:
; APPLICANT: Gully, Doris F.
; APPLICANT: Arena, Joseph P.
; APPLICANT: Liu, Ken K.
; TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
; TITLE OF INVENTION: CHANNELS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wallen, John W.
; STREET: 126 E. Lincoln Ave., P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/249,112
; FILING DATE: 25-MAY-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wallen, John W.
; REGISTRATION NUMBER: 35,403
; REFERENCE/DOCKET NUMBER: 19194

```



```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3905
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 487 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-249-112-4

alignment_scores:
  Quality: 48.00      Length: 28
  Ratio: 2.526      Gaps: 1
  Percent Similarity: 67.857      Percent Identity: 39.286

alignment_block:
US-09-684-016-48411/rev x US-08-249-112-4  ..

Align seg 1/1 to: US-08-249-112-4 from: 1 to: 487

104 AAGATAGTATCACCCTTACATGTAAGAACTTACACGTCAGTCAGCCCTCG 55
|||||  :: :::::::::::|||||:|||||
152 LysileuTySerSerArgIleSerleuThrSerSerCysPrometar 168
54 CATTCGC.....CCATTAGTCATCAGGTTGT 27
|:::|  |||||  ::|||::|::|
168 gLeuGlInleuTyrrProleuAspTyrrGlnSerCys 179

seq_name: /cgn2_6/plodata/2/1aa/PCTUS_COMB.pep:PCT-US95-06556-4

seq documentation block:
; Sequence 4, Application PC/TUS9506556
; GENERAL INFORMATION:
; APPLICANT: Cully, Doris F.
; APPLICANT: Arena, Joseph P.
; APPLICANT: Liu, Ken K.
; APPLICANT: Vassiliadis, Demetrios
; TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wallen, John W.
; STREET: 126 E. Lincoln Ave., P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06556
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/249,112
; FILING DATE: 25-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Wallen, John W.
; REGISTRATION NUMBER: 35,403
; REFERENCE/DOCKET NUMBER: 19194
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3905
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 487 amino acids
; TYPE: amino acid
```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06556-4

alignment_scores:
  Quality: 48.00      Length: 28
  Ratio: 2.526      Gaps: 1
  Percent Similarity: 67.857      Percent Identity: 39.286

alignment_block:
US-09-684-016-48411/rev x PCT-US95-06556-4  ..

Align seg 1/1 to: PCT-US95-06556-4 from: 1 to: 487

104 AAGATAGTATCACCCTTACATGTAAGAACTTACACGTCAGTCAGCCCTCG 55
|||||  :: :::::::::::|||||:|||||
152 LysileuTySerSerArgIleSerleuThrSerSerCysPrometar 168
54 CATTCGC.....CCATTAGTCATCAGGTTGT 27
|:::|  |||||  ::|||::|::|
168 gLeuGlInleuTyrrProleuAspTyrrGlnSerCys 179

seq_name: /cgn2_6/plodata/2/1aa/6B_COMB.pep:US-09-211-704A-7

seq documentation block:
; Sequence 7, Application US/09211704A
; Patent No. 6271014
; GENERAL INFORMATION:
; APPLICANT: de Saint-Vis, Blandine Marie
; APPLICANT: Fossiez, Francois
; APPLICANT: Caux, Christophe
; APPLICANT: Lebecque, Serge J.E.
; TITLE OF INVENTION: Mammalian Proteinases; Related Reagents
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/211,704A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/005,263
; FILING DATE: 09-JAN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: SF0781K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 519 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-211-704A-7
```



```

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/357,598
FILING DATE: 15-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 846 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-357-598-12

alignment_scores:
Quality: 48.00 Length: 36
Ratio: 2.400 Gaps: 2
Percent Similarity: 55.556 Percent Identity: 38.889

alignment_block:
US-09-684-016-48411/rev x US-08-357-598-12 ..
Align seg 1/1 to: US-08-357-598-12 from: 1 to: 846

105 GAGGTAAAGTATCACCCTTACATGTGAAACTTACACGATGTCAGTGCAGCCC.. 58
:::|||||:::|||||
312 ASPAASPERPHELEUHTHRCYs.....ValGlnProLe 323
57 .....TCGCATTCGGCCATTAGTGCATTCAGGCTTGTAAGCA 21
323 UGLYPRCAAPRYLYRSGLYCYSLLEUILLIETARGPROGLYPHELEUVALGLYL 340
20 TCATCAT 13
:::|||||
340 EUSERHIS 342

seq_name: /cgn2_6/prcddata/2/iaa/5B_COMB.pep:US-09-003-289-12
seq_documentation_block:
; Sequence 12, Application US/09003289
; Patent No. 5916792
; GENERAL INFORMATION:
; APPLICANT: Cytin, Curt I.
; APPLICANT: Small, Donald
; TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037

```



```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/003,289
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/357,598
FILING DATE: 15-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Haille, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 846 amino acids
TYPE: amino acid
STRANDEDNESS: No. 5916792 Relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-003-289-12
```

```
alignment_scores:
Quality: 48.00 Length: 36
Ratio: 2.400 Gaps: 2
Percent Similarity: 55.556 Percent Identity: 38.889
```

alignment\_block:

US-09-684-016-48411/rev x US-09-003-289-12 ..

Align seg 1/1 to: US-09-003-289-12 from: 1 to: 846

```
105 GAAGATAAGTATCACCCTTACATGTGAACCTTACAGTCAGTCAGCC.. 58
:::|||||: 111111111
312 AspAspSerPheLeuThrCys.....ValGlnProLe 323
57 .....TCGCATTGCCCATTTAGTTCATCAGGCTTGAAGCA 21
:::|||||: 111111111
323 uGlyProAspTyrIysGlyCysLeuIleArgProGlyPheLeuValGlyL 340
20 TCTCACAAT 13
:::|||||:
340 euserHis 342

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US95-16435-12

seq_documentation_block:
Sequence 12, Application PC/TUS9516435
GENERAL INFORMATION:
APPLICANT: The Johns Hopkins University School of Medicine
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16435
```

```
FILING DATE: 15-DEC-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Haille, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033W01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 846 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-16435-12
```

```
alignment_scores:
Quality: 48.00 Length: 36
Ratio: 2.400 Gaps: 2
Percent Similarity: 55.556 Percent Identity: 38.889
```

alignment\_block:

US-09-684-016-48411/rev x PCT-US95-16435-12 ..

Align seg 1/1 to: PCT-US95-16435-12 from: 1 to: 846

```
105 GAAGATAAGTATCACCCTTACATGTGAACCTTACAGTCAGTCAGCC.. 58
:::|||||: 111111111
312 AspAspSerPheLeuThrCys.....ValGlnProLe 323
57 .....TCGCATTGCCCATTTAGTTCATCAGGCTTGAAGCA 21
:::|||||: 111111111
323 uGlyProAspTyrIysGlyCysLeuIleArgProGlyPheLeuValGlyL 340
20 TCTCACAAT 13
:::|||||:
340 euserHis 342
```

seq\_name: /cgn2\_6/ptodata/2/1aa/GB\_COMB.pep:US-09-115-954-8

```
seq_documentation_block:
Sequence 8, Application US/09115954B
Patent No. 6200776
GENERAL INFORMATION:
APPLICANT: Boron, Walter F
APPLICANT: Brill, Antoine M
APPLICANT: Khandoudi, Nassirah
APPLICANT: Martin, Xavier
APPLICANT: Jupe, Steven C
APPLICANT: Rawlings, Christopher J
APPLICANT: Doe, Trudy R
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GH-30409
CURRENT APPLICATION NUMBER: US/09/115,954B
CURRENT FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: EP97401714.7
EARLIER FILING DATE: 1997-07-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1032
TYPE: PRT
ORGANISM: Homo sapiens
US-09-115-954-8

alignment_scores:
Quality: 48.00 Length: 12
Ratio: 4.800 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 50.000
```



## alignment\_block:

US-09-684-016-48411 x US-09-115-954-8 ..

Align seg 1/1 to: US-09-115-954-8 from: 1 to: 1032

```
5 TATTTGTAATGTCGAGATGCTTACAAACCTGATGAC 40
|||||:|||||:|||||:|||||:|||||:|||||:
623 TYRCYSARGCYSThrLeuProGluAsnProAsnAsn 634
```

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-09-115-954-2

## seq\_documentation\_block:

```
; Sequence 2, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1044
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-115-954-2
```

## alignment\_scores:

```
Quality: 48.00 Length: 12
Ratio: 4.800 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 50.000
```

## alignment\_block:

US-09-684-016-48411 x US-09-115-954-2 ..

Align seg 1/1 to: US-09-115-954-2 from: 1 to: 1044

```
5 TATTTGTAATGTCGAGATGCTTACAAACCTGATGAC 40
|||||:|||||:|||||:|||||:|||||:|||||:
635 TYRCYSARGCYSThrLeuProGluAsnProAsnAsn 646
```

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-09-352-990-18

## seq\_documentation\_block:

```
; Sequence 18, Application US/09352990
; Patent No. 6255090
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Aminoacyl-trna Synthetase
; FILE REFERENCE: BB-1191
; CURRENT APPLICATION NUMBER: US/09/352,990
; CURRENT FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,866
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Glycine max
```

US-09-352-990-18

## alignment\_scores:

```
Quality: 47.00 Length: 20
Ratio: 3.133 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 45.000
```

## alignment\_block:

US-09-684-016-48411/rev x US-09-352-990-18 ..

Align seg 1/1 to: US-09-352-990-18 from: 1 to: 400

```
104 AAGATAGTATTCACCTTACATGTGAACCTTACCACTGACAGCCCTCG 55
|||||:|||||:|||||:|||||:|||||:|||||:
4 LysMetSerValSerHisPheAlaValLeuSerSerCysCysProAr 20
54 CATTGCCCA 45
|||||:|||||:
20 glenAlaPro 23
```

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-445-586-9

## seq\_documentation\_block:

```
; Sequence 9, Application US/08445586
; Patent No. 5627050
; GENERAL INFORMATION:
; APPLICANT: Takeshita, Sunao
; APPLICANT: Ito, Toshimi
; APPLICANT: Otawara-Hamamoto, Yoko
; APPLICANT: Amano, Egon
; TITLE OF INVENTION: Bone-Related Sulfatase-Like Protein and
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,586
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/111,887
; FILING DATE: 26-AUG-1993
; APPLICATION NUMBER: JP 230030/92
; FILING DATE: 28-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324034/92
; FILING DATE: 03-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1322-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-445-586-9
```







Ratio: 3.357 Gaps: 0  
Percent Similarity: 82.353 Percent Identity: 41.176

## alignment\_block:

US-09-684-016-48411/rev x US-08-484-494-14 ..

Align seg 1/1 to: US-08-484-494-14 from: 1 to: 583

62 AGCCCTGCATTGCCCATTTAGTCATCAGGCTTGTAAAGCATCTCAT 13  
:::||||::: |||:::||||| ::||| |||:::|||||  
470 ThrpAasnpheasnpProvalglYserAsnGlYcysPhealathrhISva 486  
12 T 12  
486 I 486

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-08-345-212-14

## seq\_documentation\_block:

Sequence 14, Application US/08345212  
Patent No. 5932211  
GENERAL INFORMATION:  
APPLICANT: Willson, Peter J  
APPLICANT: Morris, Charles P  
APPLICANT: Anson, Donald S  
APPLICANT: Occhiodoro, Teresa  
APPLICANT: Bielicki, Julie  
APPLICANT: Clements, Peter R  
APPLICANT: Hopwood, John J  
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF  
TITLE OF INVENTION: IDURONATE 2-SULFATASE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/345, 212  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 991,973  
FILING DATE: 17-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8416Z  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4366  
TELEFAX: 516-742-4343  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 583 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-345-212-14

alignment\_scores:  
Quality: 47.00 Length: 17  
Ratio: 3.357 Gaps: 0  
Percent Similarity: 82.353 Percent Identity: 41.176

## alignment\_block:

US-09-684-016-48411/rev x US-08-345-212-14 ..

Align seg 1/1 to: US-08-345-212-14 from: 1 to: 583

62 AGCCCTGCATTGCCCATTTAGTCATCAGGCTTGTAAAGCATCTCAT 13  
:::||||::: |||:::||||| ::||| |||:::|||||  
470 ThrpAasnpheasnpProvalglYserAsnGlYcysPhealathrhISva 486  
12 T 12  
486 I 486

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-09-249-003-14

## seq\_documentation\_block:

Sequence 14, Application US/09249003  
Patent No. 6153188  
GENERAL INFORMATION:  
APPLICANT: Willson, Peter J  
APPLICANT: Morris, Charles P  
APPLICANT: Anson, Donald S  
APPLICANT: Occhiodoro, Teresa  
APPLICANT: Bielicki, Julie  
APPLICANT: Clements, Peter R  
APPLICANT: Hopwood, John J  
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF  
TITLE OF INVENTION: IDURONATE 2-SULFATASE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/249, 003  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/991,973  
FILING DATE: 17-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8416Z  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 583 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-249-003-14

alignment\_scores:  
Quality: 47.00 Length: 17  
Ratio: 3.357 Gaps: 0  
Percent Similarity: 82.353 Percent Identity: 41.176  
alignment\_block:







...